

OM nucleic - nucleic search, using sw model
Run on: September 30, 2003, 18:31:17 : Search time 14064 Seconds
(without alignments)
17362.739 Million cell updates/sec

Title: OS-10-019-284A-1
Perfect score: 5969
Sequence: 1 agtccgtgagccaccatt.....tgatatttggcgcgtgaa 5969

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vl.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5969	100.0	5969	6	BD093238	BD093238 DNA encod
C 2	5856.2	94.8	320550	1	AP005282	AP005282 Corynebac
C 3	5856.2	94.8	349980	6	AX127152	AX127152 Sequence
4	3687	61.8	6911	6	E11760	E11760 Base sequen
5	3687	61.8	6911	6	I26124	I26124 Sequence 4
6	1874.2	31.4	1983	6	AX122988	AX122988 Sequence
7	1874.2	31.4	1983	6	BD165105	BD165105 Novel pol
8	1457	24.4	1656	6	BD094213	BD094213 Genes for
9	1412.6	23.7	1527	6	AX069134	AX069134 Sequence
10	1412.6	23.7	1527	6	AX069134	AX069134 Sequence
11	1321.2	22.1	1342	6	AX065221	AX065221 Sequence
12	1267.8	21.2	1299	6	AX122989	AX122989 Sequence
13	1267.8	21.2	1299	6	BD165106	BD165106 Novel pol
14	1237.4	20.7	1287	6	AX065287	AX065287 Sequence
15	1237.4	20.7	1287	6	AX065289	AX065289 Sequence
16	1110.4	18.6	1152	6	AX122991	AX122991 Sequence
17	1110.4	18.6	1152	6	BD165108	BD165108 Novel pol
18	1049.8	17.6	1109	6	AX069136	AX069136 Sequence
19	762	12.8	882	6	AX065237	AX065237 Sequence
20	671.8	11.3	759	6	AX122990	AX122990 Sequence
21	671.8	11.3	759	6	BD165107	BD165107 Novel pol
22	474.2	7.9	479	6	BD094149	BD094149 Genes for
23	466.6	7.8	498	6	BD094148	BD094148 Genes for
C 24	433	7.3	343050	1	AL935252	AL935252 Lactobaci
C 25	421.6	7.1	21838	1	PDCRAFOPER	PDCRAFOPER
C 26	421.6	7.1	21839	1	PPSURFOF	PPSURFOF
27	397	6.7	5800	1	LL297015	LL297015 Lactococc
28	353.2	5.9	2508	1	STRSCRA	M22711 Streptococc
C 29	352.8	5.9	300330	1	AP005222	AP005222 Corynebac
C 30	351.6	5.9	9979	1	AE015011	AE015011 Streptoco
C 31	325	5.4	2655	1	SXSCRA	X69800 S.xylosus s
C 32	320.2	5.4	10264	1	AE004395	AE004395 Vibrio ch
C 33	320	5.4	2788	1	AF401046	AF401046 Lactobaci
C 34	318.8	5.4	301250	1	AP004830	AP004830 Staphyloc
C 35	318.2	5.3	237850	1	AP003137	AP003137 Staphyloc
C 36	318.2	5.3	341350	1	AC003385	AC003385 Staphyloc
C 37	316.6	5.3	39425	1	AC090968	AC090968 Staphyloc
C 38	303.2	5.1	10085	1	AE006222	AE006222 Pasteurel
C 39	301.2	5.0	2955	1	VIBSCRAK	M76768 Vibrio algi
C 40	293.2	4.9	66320	1	AE016833	AE016833 Enterococ
C 41	292.8	4.9	300698	1	AE016750	AE016750 Staphyloc
C 42	281.4	4.7	4158	1	BSTREAPR	254245 B.subtilis
C 43	279.4	4.7	22197	1	D83967	D83967 Bacillus su
C 44	279.4	4.7	37900	1	D86417	D86417 Bacillus su
C 45	279.4	4.7	208430	1	BS080005	Z99108 Bacillus su

ALIGNMENTS

RESULT 1
BD093238
LOCUS BD093238 5969 bp DNA linear PAT 27-AUG-2002
DEFINITION DNA encoding sucrose PTS enzyme II.
ACCESSION BD093238
VERSION BD093238.1 GI:22638826
KEYWORDS WO 0102584-A/1.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 5969)
AUTHORS Izui,M., Sugimoto,M., Nakamatsu,T. and Kurahashi,O.
TITLE DNA encoding sucrose PTS enzyme II
Patent: WO 0102584-A 1 11-JAN-2001.
JOURNAL AJINOMOTO CO INC,MASAKO IZUI,MASAKAZU SUGIMOTO,TSUYOSHI NAKAMATSU,

OSAMU KURAHASHI	OSAMU KURAHASHI	Db	721	TCCTGCTGTGCGCATTCACCTCGAGGCGCCTTTCATCAACGCATCGCGTGTGCTGCTC	780
OS	Brevibacterium lactofermentum	Qy	781	AAACCCGGATTTCATTTTCCCGGCACACCAACAGATCTTCCCGGGTGATCATCGG	840
PN	WO 0102584-A/1	Db	781	AAACCCGGATTTCATTTTCCCGGCACACCAACAGATCTTCCCGGGTGATCATCGG	840
PD	11-JAN-2001	Qy	841	GAAGAAGTTGGATCAAAATCGATCAGTAGTACGCGCGGAAACTGACAATCTTCTGAGCTTC	900
PF	30-JUN-2000 WO 2000JP004348	Db	841	GAAGAAGTTGGATCAAAATCGATCAGTAGTACGCGCGGAAACTGACAATCTTCTGAGCTTC	900
PI	02-JUL-1999 JP 99P 189512	Qy	901	TCGATCTCTCGCAGCGCACCATCATTTGCTTCCCTGGGCACACTGATGCAGATTTTG	960
PR	MASAKO IZUI,MASAKAZU SUGIMOTO,TSUYOSHI NAKAMATSU,OSAMU PI	Db	901	TCGATCTCTCGCAGCGCACCATCATTTGCTTCCCTGGGCACACTGATGCAGATTTTG	960
KURAHASHI		Qy	961	ATACCACTACACGCGCAATTGGCTTAAAGAGAAAAATGTGACGCTACGCGCTACG	1020
PC	C12N15/54,C12N9/12	Db	961	ATACCACTACACGCGCAATTGGCTTAAAGAGAAAAATGTGACGCTACGCGCTACG	1020
CC		Qy	1021	ATTGTTTCAATGGATGCTCCGCTGCATCATAGGCTCCCGGACGCTGGGCGCTTGC	1080
FT	Key	Db	1021	ATTGTTTCAATGGATGCTCCGCTGCATCATAGGCTCCCGGACGCTGGGCGCTTGC	1080
FT	CDS	Qy	1081	TTGCTGGGCGACGCTCCCGGACGCATATGTTGATGCCGACGCGGTGCATTTGG	1140
FEATURES	Location/Qualifiers	Db	1081	TTGCTGGGCGACGCTCCCGGACGCATATGTTGATGCCGACGCGGTGCATTTGG	1140
source	1..5969	Qy	1141	CCGATGGAAGCGGTGATCTAGCTCGTCCAAACAACGCCCTTTTTCATCAGGAGCGCATGG	1200
	/organism="unidentified"	Db	1141	CCGATGGAAGCGGTGATCTAGCTCGTCCAAACAACGCCCTTTTTCATCAGGAGCGCATGG	1200
	/mol_type="genomic DNA"	Qy	1141	CCGATGGAAGCGGTGATCTAGCTCGTCCAAACAACGCCCTTTTTCATCAGGAGCGCATGG	1200
	/db_xref="taxon:32644"	Db	1141	CCGATGGAAGCGGTGATCTAGCTCGTCCAAACAACGCCCTTTTTCATCAGGAGCGCATGG	1200
BASE COUNT	1388 a 1575 c 1578 g 1428 t	Qy	1201	AAGCCGCGGAAATGCCAGCGGTGATCTAGCTCGTCCAAACAACGCCCTTTTTCATCAGGAGCGCATGG	1260
ORIGIN		Db	1201	AAGCCGCGGAAATGCCAGCGGTGATCTAGCTCGTCCAAACAACGCCCTTTTTCATCAGGAGCGCATGG	1260
	Query Match 100.0%; Score 5969; DB 6; Length 5969;	Qy	1261	ATGAGTGCCTCGCTCTCGCGATGGCGGCCCATCGCCGGGGGACACAGCACACTAGCGA	1320
	Best Local Similarity 100.0%; Pred. No. 0;	Db	1261	ATGAGTGCCTCGCTCTCGCGATGGCGGCCCATCGCCGGGGGACACAGCACACTAGCGA	1320
	Matches 5969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	1321	GTGAGTTCGTGCACACAGTGGCGAGGGGTATGACGCTTATCAGCGGAGCCCTCCACACT	1380
Qy	1 AGTCGTCGACGCCACCATGATGCTGGTGCACCGAGCTTGGGAGGCTTCTACATCT	Db	1321	GTGAGTTCGTGCACACAGTGGCGAGGGGTATGACGCTTATCAGCGGAGCCCTCCACACT	1380
Db	1 AGTCGTCGACGCCACCATGATGCTGGTGCACCGAGCTTGGGAGGCTTCTACATCT	Qy	1381	CAACCGTCGCGCTAAAATTCCTGCTTGGCGATCAGCAAAATCGTAAATCCAAACCTTG	1440
Qy	61 ACCTCCCGCTCGCGTGGAGTGGGTGATACGGGTGGGATCACGCCGGTGAAGTTGG	Db	1381	CAACCGTCGCGCTAAAATTCCTGCTTGGCGATCAGCAAAATCGTAAATCCAAACCTTG	1440
Db	61 ACCTCCCGCTCGCGTGGAGTGGGTGATACGGGTGGGATCACGCCGGTGAAGTTGG	Qy	1441	CAAAATTTGTGCTTTCATCAACCGCCAGGTGCAAAAGGTCCATTTAGGTCATCAAG	1500
Qy	121 GAACCATGGTTCCTTGGTGGTGCAGGACAGTGGGTTGAGAGTTTTCACAGTG	Db	1441	CAAAATTTGTGCTTTCATCAACCGCCAGGTGCAAAAGGTCCATTTAGGTCATCAAG	1500
Db	121 GAACCATGGTTCCTTGGTGGTGCAGGACAGTGGGTTGAGAGTTTTCACAGTG	Qy	1501	TACTTTAAAGTACGAGTAAAATCTATCCTGATTTTAAAGGAGTCCCAACCAATCACTA	1560
Qy	181 TCTGCAAGTTTAAAGTATGCAATATTAACCTAGATCCGTAGACATAAGACATATA	Db	1501	TACTTTAAAGTACGAGTAAAATCTATCCTGATTTTAAAGGAGTCCCAACCAATCACTA	1560
Db	181 TCTGCAAGTTTAAAGTATGCAATATTAACCTAGATCCGTAGACATAAGACATATA	Qy	1561	TCGTGAAAGACAGCAAGTGGCGAAAGGTGCAAAAGGTCCATTTAGGTCATCAAG	1620
Qy	241 CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAAGACATATA	Db	1561	TCGTGAAAGACAGCAAGTGGCGAAAGGTGCAAAAGGTCCATTTAGGTCATCAAG	1620
Db	241 CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAAGACATATA	Qy	1621	ACAAGGTTGGAACCTTTGGGCTTGCACAGGATCCTCACCACCTGAGTACCTACCAAGAGC	1680
Qy	301 CGTCTATGCTTGTGGAAGGAACAAATTAACCTCAGAAAGATGCCAAGTGGTGAT	Db	1621	ACAAGGTTGGAACCTTTGGGCTTGCACAGGATCCTCACCACCTGAGTACCTACCAAGAGC	1680
Db	301 CGTCTATGCTTGTGGAAGGAACAAATTAACCTCAGAAAGATGCCAAGTGGTGAT	Qy	1681	TCATTCGCATGTATGAACTGGGAGGTGTCATTAAGAACTGCAAGGCACTTCTTGTGG	1740
Qy	361 ATCAAGAAATCAGGTCAGACAGTTTAAAAAATTCAGGGAAGATTTTCCCGCCCTCG	Db	1681	TCATTCGCATGTATGAACTGGGAGGTGTCATTAAGAACTGCAAGGCACTTCTTGTGG	1740
Db	361 ATCAAGAAATCAGGTCAGACAGTTTAAAAAATTCAGGGAAGATTTTCCCGCCCTCG	Qy	1741	ATGAATACCTGGGACTAACCCGCTGACGATGAAAGTAAACAGCTACTTTAAACCATTCGCAAG	1800
Qy	421 GGTGATGATGGCTTCTCCACTGAAACGGCATATCAGGAACTCTGGAGAC	Db	1741	ATGAATACCTGGGACTAACCCGCTGACGATGAAAGTAAACAGCTACTTTAAACCATTCGCAAG	1800
Db	421 GGTGATGATGGCTTCTCCACTGAAACGGCATATCAGGAACTCTGGAGAC	Qy	1801	AGTTCACTGACCATCATCGACATCGTTGATGAGAGGTCTCAGGCCACAGATGTTGCAAAAC	1860
Qy	481 CAGCACTAAAAACGCAAGATTCACCCCGAACTCCCAAGATTTTCCCGTTTATTG	Db			
Db	481 CAGCACTAAAAACGCAAGATTCACCCCGAACTCCCAAGATTTTCCCGTTTATTG	Qy			
Qy	541 ATCTTCATATCAGCGTGGAAAGCGTGGCGCTTCCCTACGGGAACGAGGACCGAGCGA	Db			
Db	541 ATCTTCATATCAGCGTGGAAAGCGTGGCGCTTCCCTACGGGAACGAGGACCGAGCGA	Qy			
Qy	601 GGAACACCGCGCATATCACCGGAAATGGCACGACCGCTGATGTTGCCAAGCATGGTT	Db			
Db	601 GGAACACCGCGCATATCACCGGAAATGGCACGACCGCTGATGTTGCCAAGCATGGTT	Qy			
Qy	661 GGGCGCGGCTGACGCACTGGCAGCGAGTGGAAACCTTATCCCTGTGTGAAGAGG	Db			
Db	661 GGGCGCGGCTGACGCACTGGCAGCGAGTGGAAACCTTATCCCTTGTGTGAAGAGG	Qy			
Qy	721 TCCTGCTGTGCGCATTCACCTCGAGGCGCCTTTCATCAACGCATCGCGTGTGCTGCTC	Db			

Db 1801 AGTTCACTGACCACATCGACATCGTTTGATGAAGAGGTCTACAGCCAGATGGTGCAAAC 1860
Qy 1861 CTGATCCATCAGAGCAGCTGCAGAGTATGAGGCAAGATCCGTCGAGATCCGTTGAG 1920
Db 1861 CTGATCCATCAGAGCAGCTGCAGAGTATGAGGCAAGATCCGTCGAGATCCGTTGAG 1920
Qy 1921 TTCAAAATCCTTGGCATCGCGGGAACGGCACATCGCTTTTCATTAACCACTATCTCTCT 1980
Db 1921 TTCAAAATCCTTGGCATCGCGGGAACGGCACATCGCTTTTCATTAACCACTATCTCTCT 1980
Qy 1981 GTGAGGACTGACAAAGGTCAGAGCTGCAGAGCTGCACCTTAACCTGAGGAGCAACGCTGAT 2040
Db 1981 GTGAGGACTGACAAAGGTCAGAGCTGCAGAGCTGCACCTTAACCTGAGGAGCAACGCTGAT 2040
Qy 2041 CTTCAACACCATCGAAGAGTCCCAACCCACGCGCTCACCCAGGTTGGGCACTTTGTC 2100
Db 2041 CTTCAACACCATCGAAGAGTCCCAACCCACGCGCTCACCCAGGTTGGGCACTTTGTC 2100
Qy 2101 CCGCGGCAAAACATCGTTGTTGTCGCACTGTGAAGGAAAGCGAGCCATCGCGG 2160
Db 2101 CCGCGGCAAAACATCGTTGTTGTCGCACTGTGAAGGAAAGCGAGCCATCGCGG 2160
Qy 2161 AACTGTGGAAGGCCAGTACTGCTTCTTCCAGGTTCCATCTGTAGATGCACAACT 2220
Db 2161 AACTGTGGAAGGCCAGTACTGCTTCTTCCAGGTTCCATCTGTAGATGCACAACT 2220
Qy 2221 GCCACCATCATCGTTGATGAAGCAGCAGTATCCAGCTGGAAACGCTGTACCTACG 2280
Db 2221 GCCACCATCATCGTTGATGAAGCAGCAGTATCCAGCTGGAAACGCTGTACCTACG 2280
Qy 2281 TCTCATGAGCAATTAAGCTCGCTAGAAACAAAAGGAAGTACTGTGGGGCTATG 2340
Db 2281 TCTCATGAGCAATTAAGCTCGCTAGAAACAAAAGGAAGTACTGTGGGGCTATG 2340
Qy 2341 CACACAGAACTTCCAGTTTGGCCCTGCGCTACCATGTGACTCCTCCGAGGGCAGGCTC 2400
Db 2341 CACACAGAACTTCCAGTTTGGCCCTGCGCTACCATGTGACTCCTCCGAGGGCAGGCTC 2400
Qy 2401 AATGATCCCAACGGAATGACGTCGATGGAGATACCTCCACGCTACTACACGACGAT 2460
Db 2401 AATGATCCCAACGGAATGACGTCGATGGAGATACCTCCACGCTACTACACGACGAT 2460
Qy 2461 CCAGGTTTCCCTTCGCACCAAGCGCACCGGCTGGCTCACACCACCGCGGTTGACC 2520
Db 2461 CCAGGTTTCCCTTCGCACCAAGCGCACCGGCTGGCTCACACCACCGCGGTTGACC 2520
Qy 2521 GGACCGCAGGATGCACTGAGCGCACCTGCCGACGCTCTTACCCGGATGATCCTAT 2580
Db 2521 GGACCGCAGGATGCACTGAGCGCACCTGCCGACGCTCTTACCCGGATGATCCTAT 2580
Qy 2581 GACCTGGATGATGCTATTCGGGTGGAGCGGTATTTACTGACGGCACACTTAACTTTTC 2640
Db 2581 GACCTGGATGATGCTATTCGGGTGGAGCGGTATTTACTGACGGCACACTTAACTTTTC 2640
Qy 2641 TACACCGGCACTTAAATGACGGAAGCGCGCCGACCCACCAACCTTTGCGAGTC 2700
Db 2641 TACACCGGCACTTAAATGACGGAAGCGCGCGCCGACCCACCAACCTTTGCGAGTC 2700
Qy 2701 GAGGACCCAACTGGGCTGATGGCGGCACTTCATCGCGGTTCCGCTTAAATTCGCTATC 2760
Db 2701 GAGGACCCAACTGGGCTGATGGCGGCACTTCATCGCGGTTCCGCTTAAATTCGCTATC 2760
Qy 2761 GACGAGCCGCGAGCGGTTTTCACACCCATTTACCGGATCCCATGATCAGCCCTGATGT 2820
Db 2761 GACGAGCCGCGAGCGGTTTTCACACCCATTTACCGGATCCCATGATCAGCCCTGATGT 2820
Qy 2821 GATGGTTGGAACATGTTCTTGGGGCCCAACGGGAAACCTCACCGGTGAGCGGTTCTA 2880
Db 2821 GATGGTTGGAACATGTTCTTGGGGCCCAACGGGAAACCTCACCGGTGAGCGGTTCTA 2880
Qy 2881 TACCGCTGCAGATCTTGAACACTGGGAATTTCTCCGGTGAATACCTTTGACCTCAGT 2940
Db 2881 TACCGCTGCAGATCTTGAACACTGGGAATTTCTCCGGTGAATACCTTTGACCTCAGT 2940

Qy 2941 GATGCACAACTGGTTCTGCTCTCTGATCTCGTTCOCGATGCTACATGTGGAAATGCCCC 3000
Db 2941 GATGCACAACTGGTTCTGCTCTCTGATCTCGTTCOCGATGCTACATGTGGAAATGCCCC 3000
Qy 3001 AACCTTTTACGCTTCGGGATGAGAAACTGGCGAAGATCTCGAGGCTGCTGATTCTTGT 3060
Db 3001 AACCTTTTACGCTTCGGGATGAGAAACTGGCGAAGATCTCGAGGCTGCTGATTCTTGT 3060
Qy 3061 CCACAAGGATTGGACCGGAATCCAGATGAGGTACTCTACTACGCAAGCTCTGACCAATGTC 3120
Db 3061 CCACAAGGATTGGACCGGAATCCAGATGAGGTACTCTACTACGCAAGCTCTGACCAATGTC 3120
Qy 3121 GGATATGCTGCGACAAGCTTGAAGAAAGCACTTCCGCGCTTTCGCGAGGATTCAGCGAG 3180
Db 3121 GGATATGCTGCGACAAGCTTGAAGAAAGCACTTCCGCGCTTTCGCGAGGATTCAGCGAG 3180
Qy 3181 CTGGGATTTCCGCGCAATCTTACGCAAGGAGGTGAGTAAACGGTCTCTGATGCGCTGG 3240
Db 3181 CTGGGATTTCCGCGCAATCTTACGCAAGGAGGTGAGTAAACGGTCTCTGATGCGCTGG 3240
Qy 3241 CTCGTGGGCTGGATGGGCTGCCCGCGAGGATGATCACCAACAGTTGCGACAGGAAGA 3300
Db 3241 CTCGTGGGCTGGATGGGCTGCCCGCGAGGATGATCACCAACAGTTGCGACAGGAAGA 3300
Qy 3301 TGGGTCGACTGCTGACTGTCGCCGCAAGCTTCAATTTCCGCAACCGCATCTACCAA 3360
Db 3301 TGGGTCGACTGCTGACTGTCGCCGCAAGCTTCAATTTCCGCAACCGCATCTACCAA 3360
Qy 3361 GAGCTCCTTCTCCAGAGGGGGAGTGGGGGTAAATCAGATCTGTATTAGGTTCTGAACCT 3420
Db 3361 GAGCTCCTTCTCCAGAGGGGGAGTGGGGGTAAATCAGATCTGTATTAGGTTCTGAACCT 3420
Qy 3421 GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTGGTCTCTGTCG 3480
Db 3421 GTCCGAGTAGACATCCGAGGCAATATTTCCCTCGAGTGGGATGGTGGTCTCTGTCG 3480
Qy 3481 GATCGTGTGTTGTCGCTGCGCTAGCTGAGGTAAACCTGGGCAATTAAGTATGATCGCGAC 3540
Db 3481 GATCGTGTGTTGTCGCTGCGCTAGCTGAGGTAAACCTGGGCAATTAAGTATGATCGCGAC 3540
Qy 3541 GATAATACAGCATTGAGATACTCAGGTGATGACAGGTTTCATTCGCTTTCCGGGC 3600
Db 3541 GATAATACAGCATTGAGATACTCAGGTGATGACAGGTTTCATTCGCTTTCCGGGC 3600
Qy 3601 CTTCAAGGTGACACTATTGAGAGATAAGTATATAAAGGGTCTTTTGTGGCGAATTTGT 3660
Db 3601 CTTCAAGGTGACACTATTGAGAGATAAGTATATAAAGGGTCTTTTGTGGCGAATTTGT 3660
Qy 3661 ACAATACTTCCGCAAAATCCCTTGATCGGACACAATAAAGAGTTTATATTTTATGCTAGC 3720
Db 3661 ACAATACTTCCGCAAAATCCCTTGATCGGACACAATAAAGAGTTTATATTTTATGCTAGC 3720
Qy 3721 TTTTGAACAAACATTCATCTCTGAATATTTTGTCTTCCCGGTTAAGGAGAAATTCAT 3780
Db 3721 TTTTGAACAAACATTCATCTCTGAATATTTTGTCTTCCCGGTTAAGGAGAAATTCAT 3780
Qy 3781 GGACCAATAAGGACCTCGCGCAACGCTGCGGACATTTGGCGGAGAGACAATTTGT 3840
Db 3781 GGACCAATAAGGACCTCGCGCAACGCTGCGGACATTTGGCGGAGAGACAATTTGT 3840
Qy 3841 CGCGCGCGCACACTGTGCAACGGTTTACGCGCTGCTCAAGACACCAAGATGTGGA 3900
Db 3841 CGCGCGCGCACACTGTGCAACGGTTTACGCGCTGCTCAAGACACCAAGATGTGGA 3900
Qy 3901 TCGCCAAAGTCTGGATGATGCCAGATCTGAAGGCACTTTTGAACCTGGCGGATGTT 3960
Db 3901 TCGCCAAAGTCTGGATGATGCCAGATCTGAAGGCACTTTTGAACCTGGCGGATGTT 3960
Qy 3961 CCAGATCATCTGCGGCCAGGCGGATGTGATCATGTTTCAAGAACTCGATGACGCAAC 4020
Db 3961 CCAGATCATCTGCGGCCAGGCGGATGTGATCATGTTTCAAGAACTCGATGACGCAAC 4020

```
QY 4021 CTCCTCAAGACATCGCTGTCTCCACAGAGAGCTCAAGATGTTGTGGCTAAACAGCCAA 4080
DB 4021 CTCCTCAAGACATCGCTGTCTCCACAGAGAGCTCAAGATGTTGTGGCTAAACAGCCAA 4080
QY 4081 CTGGTTCAGCGTGTCTGTAAGATGTTGCGGACATTTTCGTCGCGTGAATCCAACTT 4140
DB 4081 CTGGTTCAGCGTGTCTGTAAGATGTTGCGGACATTTTCGTCGCGTGAATCCAACTT 4140
QY 4141 GGTGTGGCGGTCTCTCATGGCTATCAACATGTTGGTGTGGCAGGATCTGTGG 4200
DB 4141 GGTGTGGCGGTCTCTCATGGCTATCAACATGTTGGTGTGGCAGGATCTGTGG 4200
QY 4201 TCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTGTGTGAGATGATCAAC 4260
DB 4201 TCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTGTGTGAGATGATCAAC 4260
QY 4261 GATGCGATCTGCGCCCTTCGCGTCTCTGCAAGTGTGTGGTGTGGCAGGATCTGTGG 4320
DB 4261 GATGCGATCTGCGCCCTTCGCGTCTCTGCAAGTGTGTGGTGTGGCAGGATCTGTGG 4320
QY 4321 TTTCGGTGGCAATGAGTTCCTGGCGCCGGCATTTGGTATGGGATGGTGTTCGCCAAGC 4380
DB 4321 TTTCGGTGGCAATGAGTTCCTGGCGCCGGCATTTGGTATGGGATGGTGTTCGCCAAGC 4380
QY 4381 GGTAAACGGTACAGCTGCGCCACCATGACCGCGCGGAAATGCCAATGTGGTCCCT 4440
DB 4381 GGTAAACGGTACAGCTGCGCCACCATGACCGCGCGGAAATGCCAATGTGGTCCCT 4440
QY 4441 GTTTGGTTGGATGTTGCTCAAGCTGGTTACAGGGCACCGTCTCTGTGCTGGTGT 4500
DB 4441 GTTTGGTTGGATGTTGCTCAAGCTGGTTACAGGGCACCGTCTCTGTGCTGGTGT 4500
QY 4501 CTCTTGATTTCTGGCAACGATCAGAAATTCCTGCACAAAGCACTCATGGGACATGCAGA 4560
DB 4501 CTCTTGATTTCTGGCAACGATCAGAAATTCCTGCACAAAGCACTCATGGGACATGCAGA 4560
QY 4561 CTTCCTGATCACCAGTGTGACTCTGCTGCTCACCAGGCTTCCCTACGTTGATTCCTAT 4620
DB 4561 CTTCCTGATCACCAGTGTGACTCTGCTGCTCACCAGGCTTCCCTACGTTGATTCCTAT 4620
QY 4621 TGTCTCAGCAATCGGTGGGTGACTTGTGGCACACGGTCTGCAGGGACTCTATGA 4680
DB 4621 TGTCTCAGCAATCGGTGGGTGACTTGTGGCACACGGTCTGCAGGGACTCTATGA 4680
QY 4681 TTTCGGTGGTCCAGTGGCGGTGTGCTTTTCGGTGTGGTCTACTACCAATCGTTATCAC 4740
DB 4681 TTTCGGTGGTCCAGTGGCGGTGTGCTTTTCGGTGTGGTCTACTACCAATCGTTATCAC 4740
QY 4741 TGTCTCAGCAATCGGTGGGTGACTTGTGGCACACGGTCTGCAGGGACTCTATGA 4800
DB 4741 TGTCTCAGCAATCGGTGGGTGACTTGTGGCACACGGTCTGCAGGGACTCTATGA 4800
QY 4801 CATCTTTCGCAACCGCATCCATGCCAATATCGCGCAGGGTGCAGCATGTTTGGCAGTGT 4860
DB 4801 CATCTTTCGCAACCGCATCCATGCCAATATCGCGCAGGGTGCAGCATGTTTGGCAGTGT 4860
QY 4861 CTTCCTAGCAAGAGTGAAGCTCAAGGCTCAGGCTTGCAGGTGTCTCCGCTGT 4920
DB 4861 CTTCCTAGCAAGAGTGAAGCTCAAGGCTCAGGCTTGCAGGTGTCTCCGCTGT 4920
QY 4921 TCTTGTGATTACAGAGCTCGATCTTCGGTGTGAACCTTTCGCTCGCTCGCTGCTTCTA 4980
DB 4921 TCTTGTGATTACAGAGCTCGATCTTCGGTGTGAACCTTTCGCTCGCTCGCTGCTTCTA 4980
QY 4981 CATGTGATCGGTACCGCAGCTATCGGTGGCGCTTTGATGTGACTCTTTGATATCAAGC 5040
DB 4981 CATGTGATCGGTACCGCAGCTATCGGTGGCGCTTTGATGTGACTCTTTGATATCAAGC 5040
QY 5041 AGTTGCGTTGGCGCTGCAGGTTCTTGGTGTGTTCTTATGATGCTCCAGATATGT 5100
DB 5041 AGTTGCGTTGGCGCTGCAGGTTCTTGGTGTGTTCTTATGATGCTCCAGATATGT 5100
QY 5101 CATGTTCTTGGTTGGCGGTAGTACCTTTGTTCATCGCAATTCGGCGCAGCGATTCCTTA 5160
```

```
DB 5101 CATGTTCTTGGTTGGCGGTAGTACCTTTGTTCATCGCAATTCGGCGCAGCGATTCCTTA 5160
QY 5161 TGGCTTTTACTGTTGTTCCGCGCACGACGATTCAGATTCACACCGTGTCTCCAGT 5220
DB 5161 TGGCTTTTACTGTTGTTCCGCGCACGACGATTCAGATTCACACCGTGTCTCCAGT 5220
QY 5221 GCCTGCAGGAACGACCAAGCCGAGAGCAGACCCGCGAGATTTTCAACAGATTCAC 5280
DB 5221 GCCTGCAGGAACGACCAAGCCGAGAGCAGACCCGCGAGATTTTCAACAGATTCAC 5280
QY 5281 CATCATCCAGCAGCTTTGACCGGTGAGCTATCGCACTGAGCAGCGTTCAGCATGCGAT 5340
DB 5281 CATCATCCAGCAGCTTTGACCGGTGAGCTATCGCACTGAGCAGCGTTCAGCATGCGAT 5340
QY 5341 GTTTGCCAGCGAAAGCTTGGCTCAGGTGTTCGATCCTCCCCACCAAGGGGAGCTGGT 5400
DB 5341 GTTTGCCAGCGAAAGCTTGGCTCAGGTGTTCGATCCTCCCCACCAAGGGGAGCTGGT 5400
QY 5401 TTCACCACTGAGCGGAAGATCGTGGCTTCCCATCTGCTCAGCTTTCGAGTCCG 5460
DB 5401 TTCACCACTGAGCGGAAGATCGTGGCTTCCCATCTGCTCAGCTTTCGAGTCCG 5460
QY 5461 CACTAAGCTGAGGATGGTTCCTCAATGTGGATATCTTGATGCACATTTGTTTCGACACCGT 5520
DB 5461 CACTAAGCTGAGGATGGTTCCTCAATGTGGATATCTTGATGCACATTTGTTTCGACACCGT 5520
QY 5521 AAACCTCAACGCGCAGCACTTTAACCGCTGAAGAACGAGCGCATGAAGTCAAGCAGG 5580
DB 5521 AAACCTCAACGCGCAGCACTTTAACCGCTGAAGAACGAGCGCATGAAGTCAAGCAGG 5580
QY 5581 GGAGTGTGTGTGAATTCGATATTTGATGCCATTAAGCTGCAGGTTATGAGGTAACAC 5640
DB 5581 GGAGTGTGTGTGAATTCGATATTTGATGCCATTAAGCTGCAGGTTATGAGGTAACAC 5640
QY 5641 GCCGATTTGTTTTCGAATTAACAAGAACCGCACTTGAACACTTACGGTTTGGCGGA 5700
DB 5641 GCCGATTTGTTTTCGAATTAACAAGAACCGCACTTGAACACTTACGGTTTGGCGGA 5700
QY 5701 AATTGAAGCGGAGCCACCTGCTCAACGTCGAAAGAAAGAGCGGTGCCAGCAACACC 5760
DB 5701 AATTGAAGCGGAGCCACCTGCTCAACGTCGAAAGAAAGAGCGGTGCCAGCAACACC 5760
QY 5761 ATAGTTGAACCTTGAGTGTTCGCACACAGGTTAGACTAGGACGTTGACTCTACGAT 5820
DB 5761 ATAGTTGAACCTTGAGTGTTCGCACACAGGTTAGACTAGGACGTTGACTCTACGAT 5820
QY 5821 CTTTCACACCGGTACCCGTACGCTTCGAGATTTTAAACCTGTTCAACCAAGTTCATGCTC 5880
DB 5821 CTTTCACACCGGTACCCGTACGCTTCGAGATTTTAAACCTGTTCAACCAAGTTCATGCTC 5880
QY 5881 GGTGTACCTGTGTGTGCCACCCCGCAATCTTCACCCCAACATTTGGACATGTTCTGCTCAGC 5940
DB 5881 GGTGTACCTGTGTGTGCCACCCCGCAATCTTCACCCCAACATTTGGACATGTTCTGCTCAGC 5940
QY 5941 AGTAGCGTTGATATTTTGGCGCGCTGAA 5969
DB 5941 AGTAGCGTTGATATTTTGGCGCGCTGAA 5969
```

RESULT 2

AP005282/c

LOCUS

DEFINITION Corynebacterium glutamicum ATCC 13032 DNA, linear BCT 08-AUG-2002

AP005282 320550 bp

AP005282 BA000036

AP005282.1 GI:21325287

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Corynebacterium glutamicum ATCC 13032

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE 1

AUTHORS Nakagawa,S.
 TITLE Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 320550)
 AUTHORS Nakagawa,S.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co.
 Ltd., Tokyo Research Laboratories: 3-6-6, Asahi-nachi, Machida,
 Tokyo 194-8533, Japan (E-mail:snakagawa@xanagen.com,
 Tel:81-44-829-3031, Fax:81-44-813-1651)
 COMMENT This sequence is conducted by collaboration of Kyowa Hakko Kogyo
 Co. Ltd. And Kitasato University.
 FEATURES
 Location/Qualifiers
 1. 320550
 /organism="Corynebacterium glutamicum ATCC 13032"
 /mol_type="genomic DNA"
 /strain="ATCC 13032"
 /db_xref="taxon:196627"
 /note="ATCC 13032"
 188..814
 /gene="Cg12516"
 188..814
 /gene="Cg12516"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical protein"
 /protein_id="BAB99909.1"
 /db_xref="GI:21325288"
 /translation="MYRAFDVPLSMVLVTKTRKDRMETLAAQARTLLEKGVAP
 THAFVSIKAIPILSILLITLVTVNGISGNVQPPALEQVRTDVYKININERLK
 GLVSPLELHTAQTAKRNADSSEKVPDEGLVQLQNLPIYANADTIYDR
 FLSPDHVKLLLDYEAIGVGAYKGDHAWIVVFTVAPADSVESTE"
 811..2151
 /gene="Cg12517"
 811..2151
 /gene="Cg12517"
 /codon_start=1
 /transl_table=11
 /product="Nicotinic acid phosphoribosyltransferase"
 /protein_id="BAB99910.1"
 /db_xref="GI:21325289"
 /translation="MNTNPFSSNNRSTALLTDKYELTMQAAADGASRPSTFEV
 SRRLNRRYGVAGTARVLAIRDVFTEQLADLDLDRTLEYLRNYRFTGVQV
 YREGELVPSPLTVRGTFACVILETVILSINNADSAVASAARVTAADGRPIIE
 MGRSTHYSAYTSARAYLAGFTSTNSLEARYRGIPASGTSAAHWTLIHNDDGP
 NEAAKFAVESLGVDTLLVDYDITCGVATALEVAGPDGLGVRIDSGLVARV
 RKQDLDNAHTKIVYSSDLDEFAIAGLRGPDVGVGTSVVTCGAPTAGLVYKIV
 EVAGHPVAKRNKESYGGKKAVRTHKSGTAIEEIVYFENAEAPDTGKLDLTSLI
 PLMRDGETVGPGLPLEDSRAYLAKQLVSLPWEGLASNDPEVLHTRFVGPFPAA"
 2438..2746
 /gene="Cg12518"
 2438..2746
 /gene="Cg12518"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical protein"
 /protein_id="BAB99911.1"
 /db_xref="GI:21325290"
 /translation="MGLRNLFGRPPOKTLQTKLPAQRNPQAPLSKOSNOTVNRCEA
 WLSLSTPWLQAOCLRLNHPQIPQIRSMPLPRKSCILLPLKRWAVPVLDKQWLRL
 2866..4608
 /gene="Cg12519"
 2866..4608
 /gene="Cg12519"
 /codon_start=1
 /transl_table=11
 /product="Rad3-related DNA helicases"
 /protein_id="BAB99912.1"
 /db_xref="GI:21325291"
 /translation="MSTATIALQROLVNRDLPLVDALEPLMERRPFAIMKGRSNYL
 CMNKVAROEELNQEDLSWLGKHIVRLNEWANETETGDRDLDPQVPLANK
 QVSVTARECIGASRCPHGDCFAETARGKAEADVYVVTNHALLDALSDVSVLPEDH

VVVIDAEHLDGRIITAVASAEITVNSLNAARRASKLDSKREERVQRTAGDLETLQ
 TMQGRNMDSGSKTGLVALKDALWALRAQIAGAPEGAANDPERFAERQNLNHLM
 EHDANRLLEVFAEEDPSKYDYVWNHDDRRGSLNAPLSVAGLLREKLFASNTY
 VLASATLTIGGNFNAWASWGLPKGSWSDMDAGTFFPAKSLITARHLKPPDGDGL
 PEETLDEIYELITAAAGRTLGLFSKRAAEAKTAKMLRDPDFVLCQDDNTAAALYKK
 FDSSENTCLFTLTLQWGVDFGSRSLVIDRIPFPDPDPLLLQARKEADAEGNRG
 FVEVAAHAALLMAQAGRLRHVGRGVAVVLDHRLSTKRYGGFLRFSPRFWETTN
 PETVRAALKRLVTK"
 complement(4618..5511)
 /gene="Cg12520"
 complement(4618..5511)
 /gene="Cg12520"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical membrane protein"
 /protein_id="BAB99913.1"
 /db_xref="GI:21325292"
 /translation="MTSPGSGQVPGDDNNWNWSQFNGPSGEQNGQPYGAPYGPYGVQ
 PDQGFNAYSPPIPEVPQPSMQEAOQRSFDLGTGFGQAGKGTATQWAWLSALIYF
 AVLLVLMFAWILPMVSLAATSGSDSAIAATGTSFFEGMLMIVLAFISVYSLNC
 YRNAARVWREGIITIOSFFKMGKLGALGIYILINIVIFIGMILLIPGIIAAVLIIF
 AVPAFQLRCASIGDASFAFSKAVSKVNGVQVILLELAIFALSGSAVIGMLVTTPL
 TELLYAYAQTSAGSGPIMQRQ"
 complement(5587..6312)
 /gene="Cg12521"
 complement(5587..6312)
 /gene="Cg12521"
 /note="PF01981:Domain of unknown function UPF0099"
 /codon_start=1
 /transl_table=11
 /product="Uncharacterized ACR"
 /protein_id="BAB99914.1"
 /db_xref="GI:21325293"
 /translation="MPKDSLLDAHLLLKARSDGPHSDPSDPSTVOAMQIALHIPKQ
 QDLPGVTVEDDAAKVRAFLPSADVDPADIKKLOISGTPLDPEPNINDEYPIVID
 ESLKMTLGKAAQVGHASKMLAAHQFFENVEQWAEADFAHVREIPSEEFRLIESPG
 AVFVRGGTEVAPNTVVAIP"
 complement(6305..7324)
 /gene="Cg12522"
 complement(6305..7324)
 /gene="Cg12522"
 /EC_number="3.1.3.3"
 /note="PF00702:haloacid dehalogenase-like hydrolase
 TIGR00099; TIGR00099; conserved hypothetical protein
 TIGR00338;serf: phosphoserine phosphatase Serf"
 /codon_start=1
 /transl_table=11
 /product="Phosphoserine phosphatase"
 /protein_id="BAB99915.1"
 /db_xref="GI:21325294"
 /translation="MHGQSVVVELQETVQSSRRSSHVVVLGDPVDALDISRIGOTL
 ADYDANIDIRIGSDPYVGLKLVTPVPSGCGEAMKALALSELNVDIAIERS
 GLLRKRLVCFDCDSTLITIGTEIEMLAHAGREAVAVTERAMRGELDPELSLER
 KALAGLDSVIDEAAALTEPTGATTINTLRMGYQTAVSGGQIQVLEGLAELE
 LDYVRANTLEIVDGKLTGNVTIKVDRAAKAEFLREFAADSGLKMYQTAVAGDGANDI
 DMSAAGLGVAFNAKPALEIADTSVNHVFLDEVLHIMGISDEIDLADQEDGTFRV
 PLTNA"
 complement(7811..9565)
 /gene="Cg12523"
 complement(7811..9565)
 /gene="Cg12523"
 /EC_number="1.9.3.1"
 /note="PF00115:Cytochrome C and Quinol oxidase polypeptide
 I"
 /codon_start=1
 /transl_table=11
 /product="Heme/copper-type cytochrome/quinol oxidases,
 subunit 1"
 /protein_id="BAB99916.1"
 /db_xref="GI:21325295"
 /translation="MTAVAPRVGDHVAQPERPEPTGHARGKSAWLMTTTTHKQLGIM
 YIIMSFSEFFLGLMALLIRAEFLTQQLFSLNQFNQLFTMGTGTVALLLYGTPIVMG

gene	FANYVLPAGADVAFPRNLNFAFWITTVGVAMLTGLTFPGAADFQWMTYSPUSD	
	AIRSPGLGDMWTVGVGATIGSVASAINMLTILCLRAPMTMFMPIFTWNIFFVFS	
CDS	VLALLFPLLLAALGVLDKRLGHLDPANGSLWQHLFWFFGHFVYVIALPFF	
	GIYSELIIPFSRPMGYVGLIFATLSIGALSMAWHAHMEVTLGAVLLPFFSFMFLI	
	SVPTGVKFNWNGTGMKGHTWETPMIWSVGMFAPELFGGLTGIMLSPDLPHLADS	
	YELLAHFHTLFTGYVASCAGYVFWFPMWTGRMDERLCKHFNLTFVFGFTGLIQ	
	HWYNGMFRATYDSDGFIYNOISIVFSFLGLSVIPFIWNVKRWKRGELTV	
	DDPWGNSLEWATWISCPFPRPHNFASLPRIISRPAFELPHMIERMRAEHTGHDD	
gene	INAPELGTAPALASDSR"	
	9701..10009	
CDS	/gene="Cg12524"	
	9701..10009	
	/gene="Cg12524"	
	/codon_start=-1	
	/transl_table=11	
	/product="Hypothetical protein"	
	/protein_id="BAB99917.1"	
	/db_xref="GI:21325296"	
gene	/translation="MPKKNVFOIHEVAAREFSLWAVTENQCEGVHNLTDISPSPT	
	PTVNPCHRPNSPRRSIEWLRPFQGRFLKXEDVPRGRGAWRSPTICASERPA	
CDS	complement(10086..11090)	
	/gene="Cg12525"	
	complement(10086..11090)	
	/gene="Cg12525"	
	/note="PF00268:Ribonucleotide reductase, small chain"	
	/codon_start=-1	
	/transl_table=11	
	/product="Ribonucleotide reductase beta subunit"	
	/protein_id="BAB99918.1"	
	/db_xref="GI:21325297"	
gene	/translation="MAADSLVHDAYLKEHVAPVKAINWNSIPDSKDLEVWDRLTGN	
	FWLPKVPVNSDKSWGTNEVEKAATMRVFTGLTLDITQGTVGAISSLPPADSLSHE	
CDS	EAVLTNIAFNSVHAKSYSNIFMTLASTABIANDAFWSEENENLORAKIILSYVEGD	
	DPLKRTIASVILEFYSGLFPLPMYSSHSKLTNTADVRLTIIRDEAVHGYIYGVY	
	OKAVAKETPERQELKEYFDLLYLDYNETQXSEDLYYDLSWTEVYKRELYNANKA	
	LNKLGEGLFPADETKVPNLSALSPNADENHDFYSGSGSSYVIGKAENTDEDDWDF	
gene	11555..12043	
	/gene="Cg12526"	
CDS	11555..12043	
	/gene="Cg12526"	
	/note="PF00210:Ferritin	
Query Match 94.8%; Score 5656.2; DB 1; Length 320550;		
Best Local Similarity 97.1%; Pred. No. 0;		
Matches 5793; Conservative 0; Mismatches 173; Indels 3; Caps 3;		
Qy	1	AGTCCGTCGAGCCACCATGATGCTGGTGTCAACGAGCTTCGGAGGCTTTCTACATCT 60
Db	152336	
Qy	61	ACGCTCCCGTCGGCGTGGAGTGCGTCAATACGGGTGGGATCACGCCGGTGAAGTTGG 120
Db	152276	
Qy	121	GAACCCATGGTGTCTCTGGTGTAGGGAAGAGTTCGGGTGAGAGTCTTTCAGTG 180
Db	152216	
Qy	181	TCTCAGTCTTTAAGTTATGCATCATCAGCTTGAAGGCTGAGGCTGAGAGTCTTTCAGTG 152157
Db	152156	
Qy	241	CAACAGCAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAGACATCAT 300
Db	152096	
Qy	301	CGTCCCTATGTTGCTGGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTGCAAT 360
Db	152036	
Qy	1441	GAATTTTGTGTCCTTGNACTCAACGCCAGGTGCAGAAAGTTCATAGGTCCATCAAG 1500

Db	151976	
Qy	421	GGTGATTTGATGGCTTTCTCCAACTCGAAACGCGCATCATCAGGAACCTCTCTGGAGAAC 480
Db	151916	
Qy	481	CAGCACCTAAAAACGACGAGATTCACCCCGCAACTCCCCACGATTTGCCCGGTTTATTG 540
Db	151856	
Qy	541	ATCTTCATAATCAGCGTGGAAACGCTGGCGGTTTCTTACGGGAACGAGCAGCAGCGGA 600
Db	151796	
Qy	601	GGAACCCGCGGAGTATCACCCGCGAAATGCGACACGCGTGATGTTGCCAAGCATGTTT 660
Db	151736	
Qy	661	CGCGCCGCGCTGACCACTGGCAGCGAGGTGGAAACCTTATCCCTTGTGTGAAGAGG 720
Db	151676	
Qy	721	TCTGCTGTGGCGCATCACCTCGAGGCCCTTTTATCAACGCGATCGCGTTGTGGTGTC 780
Db	151616	
Qy	781	AAAAACCGGATTTCAATTTTCCCGCAACCCCAACAGATCTTCCCGGGTGATCCATCGG 840
Db	151556	
Qy	841	GAAGAGTTGGATCAATTCGATCACAGTAGCGCGCGGAACATGACAATCTTTTACGCTTC 900
Db	151496	
Qy	901	TGATCTCTGCCAGCGGCACACATCATGTCTCTTCCGGGCACACTGATCAGATTTTG 960
Db	151436	
Qy	961	ATACCACTACCAAGCGCAATTTGCTTGGCTAAAGAAAAATGTGACGGTCAACGGCTACGC 1020
Db	151376	
Qy	1021	ATTGTTCAATCGATGCTTCGCTGCATCATAGGGTCCCGGAGCGTGGCGGCTTTGC 1080
Db	151316	
Qy	1081	TTGCTGGSCACGTGCCGGGACGCATATGTTGAGTTGATCGCGACGCGGTGCAATTTGG 1140
Db	151256	
Qy	1141	CCGATGGAAACGCTGATAGCTCGTCCAAACAGCCCTTTTTCATCAGGACGCCATGG 1200
Db	151196	
Qy	1201	AAAGCCGCGGAATGCCAGCGGTGAGTACATTTTGGCGTGTGAACTGACCGTCAACGTCACG 1260
Db	151136	
Qy	1261	ATGGAGTCCCGCTCTGGCGGATGGCGGCGCATCGCGGGGGGACACAGCACTAGGCGA 1320
Db	151076	
Qy	1321	GTCACTTCGTGCACACGCTGCGCAGGGTATGACCGTTATCGACGCGACCTTCCACACT 1380
Db	151016	
Qy	1381	CAACGCTCCCGCTAAAAATCTCGGTCTTGGCGATCAGGAAATCGCTAAATCCAACTG 1440
Db	150956	
Qy	1441	GAATTTTGTGTCCTTGNACTCAACGCCAGGTGCAGAAAGTTCATAGGTCCATCAAG 1500

150896	TAAATTTTGTGGTCTTTTGACTCAAAACGGCCAGTTTACACAGGTCCTCATTTAGACCACTATCAAG	150837
1501	TACTTTAAGTACGAGTAAATACTATCTCTGAT-TTTAAAGGAGTCCCAACCATGGAATCACT	1559
150836	TAATTTAAATACGAGCAAAACTTTTCTGTATATAATAAAGGAGTCGACCATGACATCATC	150777
1560	ATCTGCAAGACGAGCAAGAGTCGCAAGACGATTGCGAGTCTTAATCGCACCCCTCGCC	1619
150776	ATCTGCAAGACGAGCAAGAGTCGCAAGACGAGCGCGCCCTGATCGACCCCTTCGCA	150717
1620	AACAAGGTTGGAACCTTTGGGGCTTGCAACAGGATCTCCACCATGAGTACCTACCAAGAG	1679
150716	ACTAAGGGCGGAACCTTTGGGGCTTGCAACTGGATCTGCACCTTTGAGCACCTACCAAGAG	150657
1680	CTCATTCGCATGATGAAGCTGGGAAAGTGTCAATTCAGAACTGCAAGGATCTCTTGTTG	1739
150656	CTCATTCGCATGATGAAGCTGGGAAAGTGTCAATTCAGAACTGCAAGGATCTCTTGTTG	150597
1740	GATGAATACGTGGGACTTAACCCGTGACGATGAACAGGATCTTTAAACCATTCGCAAA	1799
150596	GATGAATACGTGGGATTAACGCGCGAGCATGAACAGTACTTCAAACCATTCGTAAA	150537
1800	GAGTTCACTGACCACTCGACATCGTTGTATGAAGAGTGTACAGCCCGCATGGTGCAAAC	1859
150536	GAGTTCACTGACCACTCGACATCGTTGTATGAAGAGTGTACAGCCCGCATGGTGCAAAC	150477
1860	CCTGATCCATACGAAGCAGTGCAGAGTATGAGGCAAAAGATCGCTGCAGAACTCCGTTGAA	1919
150476	CCTGATCCATACGAGCAGTGCAGAGTATGAGGCAAAAGATCGCTGCAGAACTCCGTTGAT	150417
1920	GTTCAAAATCCCTTGCAATCGGCGGAAACGG-CACATCGCTTTCAATGAACCATCATCTTCT	1978
150416	GTTCAAAATCCCTTGCAATCGGCGGAAACGGCGCATCGCTTTCAATGAGCATCATCTTCT	150357
1979	CTGTACGAGTACAAAGTTCACAGGCGTCCACCCCTAAAACCTGTGGAGACAAACGCTCGA	2038
150356	CTGTACGAGTACAAAGTTCACAGGCGTCCACCCCTAAAACCTGTGGAGACAAACGCTCGA	150297
2039	TTCTTCAACACCATCGAAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGGCACTTTG	2098
150296	TTCTTCAACACCATCGAAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGGCACTTTG	150237
2099	TCCCGCGCGCAAAACATCGTTGGTGGCACTGTGGTGAAGAAAGCCGACGCCATCCGC	2158
150236	TCCCGCGCGCAAAACATCGTTGGTGGCACTGTGGTGAAGAAAGCCGACGCCATCCGC	150177
2159	GGAAGTGTGGAAGCCCGAGTACGTCTTCTTGCCAGGTTCCATCTCTGATAGTGCACAAAC	2218
150176	GGAAGTGTGGAAGCCCGACATGACCGCATGTGCCAGGTTCCATCTCTGAGATGCAACAAAC	150117
2219	ATGCCACCATCATCGTTGGATGAAGCAGCAGTATCCAAGCTGGAACAAACGCTGATCACTAC	2278
150116	AATGCCACCATCATCGTTGGATGAAGCAGCAGTATCCAAGCTGGAACAAACGCTGATCACTAC	150057
2279	CGTCTCATGGAGCAATTAAGAGCTCGCGTAGAAAACAAAAGAAAGTACTGTGTGGGGCTA	2338
150056	CGTCTCATGGAGCAATTAAGAGCTCGCGTAGAAAACAAAAGAAAGTACTGTGTGGGGCTA	149997
2339	TGCACACAGAACTTTCAGTTTGGCCCTGCGTACCATGTGACTCCTCCGACGGGACGC	2398
149996	TGCACACAGAACTTTCAGTTTGGCCCTGCGTACCATGTGACTCCTCCGACGGGACGC	149937
2399	TCAATGATCCCAAGGAATGTACGTGATGGAGATACCCCTCCACGTCTACTACCAAGCAGC	2458
149936	TCAATGATCCCAAGGAATGTACGTGATGGAGATACCCCTCCACGTCTACTACCAAGCAGC	149877
2459	ATCCAGGTTTCCCTTCGCAACAAAGCGCACCGGCTGGGCTGCACACCAACCGCCTTGA	2518
149876	ATCCAGGTTTCCCTTCGCAACAAAGCGCACCGGCTGGGCTGCACACCAACCGCCTTGA	149817
2519	CCGACCCGACGATTCGAGTGGAGCAGCACTGCCGACGCTCTTTACCCGGATGCACTCT	2578
149816	CCGACCCGACGATTCGAGTGGAGCAGCACTGCCGACGCTCTTTACCCGGATGCACTCT	149757

Qy	3659	GTACAAATACCTCGCAAAATCCCTTGATCGGACACAAATAAACAAGTTTAAATATGTTTA	3718		Db	147597	ACTGCTCGCACCACTCTCCCGCAATGAGCTGGAGCTGTGTTAACCGAGGTGGATCC	147538	
Db	148677	GTACAAATACCTCGCAAAATCCCTTGATCGGACACAAATAAACAAGTTTAAATATGTTTA	148618		Qy	4799	TTCACTCTCGCAACCGCATCCATGCGCAATATCGCGAGGTCAGCATGTTTGGCAGTG	4858	
Qy	3719	GCTTTTGAACAAACATTCATGCTGAATATTTTGTCTCTCCCGTTAAGGAAATTC	3778		Db	147537	TTCACTCTCGCAACCGCATCTATGCTAATATCGCCAGGTCGGGCATGTTTGGCAGTG	147478	
Db	148617	GCTTTTGAACAAACATTCATGCTGAATATTTTGTCTCTCCCGTTAAGGAAATTC	148558		Qy	4859	TTCTTCTTAGGAAAGAGTGAAGAGCTCAAGGCGCTTGAGGTCGTTCAAGTGTCTCCGCT	4918	
Qy	3779	ATGACCATTAAGGACCTCGCGCAACGATCCTCGGGACATTTGGCGCGCAAGACACAT	3838		Db	147477	TTCTTCTCGGCAAGAGTGAAGAGCTCAAGGCGCTTGAGGTCGTTCAAGTGTCTCCGCT	147418	
Db	148557	ATGACCATTAAGGACCTCGCGCAACGATCCTCGGGACATTTGGCGCGCAAGACACAT	148498		Qy	4919	GTCTTGTGATTTACAGAGCTCGCATCTTGGTGTGAACCTTCGCTCGCTCGCTGGCGCTTC	4978	
Qy	3839	GTGCGCGCGCACACTGTGCAACGCGTTTACGCGCTGTGCTCAAAACACACACAGATG	3898		Db	147417	GTCTTGTGATTTACAGAGCTCGCATCTTGGTGTGAACCTTCGCTCGCTGGCGCTTC	147358	
Db	148497	GTGCGCGCGCACACTGTGCAACGCGTTTACGCGCTGTGCTCAAAACACACACAGATG	148438		Qy	4979	TACATTTGTATCGGTACCGACGCTATCGGTGGCGCTTTGATGACACTCTTTTGAATCAAG	5038	
Qy	3899	GATCGCAAAAGTCTGGATGATGATCAGATCTGAAAGGACCTTTTGAAGTCTGCGCATG	3958		Db	147357	TTCACTCGGTATCGGTACCGACGCTATCGGTGGCGCTTTGATGACACTCTTTAATATCAAG	147298	
Db	148437	GATCGCAAAAGTCTGGATGATGATCAGATCTGAAAGGACGTTTGAAGCGGTGTATG	148378		Qy	5039	CGAGTGTGCTGGCGCTGCAAGTTTCTTGGTGTGTTTCTTATGATGCTCCAGATATG	5098	
Qy	3959	TTCCAGATCATCTGCGCGCGCAGCGGATGAGTATCATGTTTCAAGAACTCGATGACGA	4018		Db	147297	CGAGTGTGCTGGCGCTGCAAGTTTCTTGGTGTGTTTCTTATGATGCTCCAGATATG	147238	
Db	148377	TTCCAGATCATCTGCGCGCGCAGCGGATGAGTATCATGTTTCAAGAACTCGATGACGA	148318		Qy	5099	GTCACTGTTCTTGGTTTGGCGGTAGTTTACCTTTGTCTATCGCATTCGGCGCAGCGATGCT	5158	
Qy	4019	ACCTCCAAAGACATCTGCTGTGTCACAGACGATCTCAAGATGTTGTGGTAAACAACGCC	4078		Db	147237	GTCACTGTTCTTGGTGTGCAAGTTTCTTGGTGTGTTTCTTATGATGCTCCAGATATG	147178	
Db	148317	ACCTCCAAAGACATCTGCTGTGTCACAGACGATCTCAAGATGTTGTGGTAAACAACGCC	148258		Qy	5159	TAAGCGCTTACTTGGTTCGCGCAACCGCACCATTTGATCAGATGCAACCGCTGCTCCA	5218	
Qy	4079	AACGTGTTACGCGTCTGTAAGATTTTGGCGGACATTTTCCTCCCGTGTATTCGAATC	4138		Db	147177	TAAGCGCTTACTTGGTTCGCGCAACCGCACCATTTGATCAGATGCAACCGCTGCTCCA	147118	
Db	148257	AACGTGTTACGCGTCTGTAAGATTTTGGCGGACATTTTCCTCCCGTGTATTCGAATC	148198		Qy	5219	GTGCTGTGAGGAACACCAAGCCAGCAAGCAACCGCGCAGATTTTCAAGCAATTC	5278	
Qy	4139	TTGTTTGTGCGGTCTGCTCATGCTATCAAAATGTTGTTGTTGGCAGGATCTGTTTC	4198		Db	147117	GTGCTGTGAGGAACACCAAGCCAGCAAGCAACCGCGCAGATTTTCAAGCAATTC	147058	
Db	148197	TTGTTTGTGCGGTCTGCTCATGCTATCAAAATGTTGTTGTTGGCAGGATCTGTTTC	148138		Qy	5279	ACCATCATCCAGGACCTTTTGAACCGTGAAGTATCGCACTGAGCAGCGTCAGCGATGCC	5338	
Qy	4199	GGTCCGCAATCACTGTGAGATGTTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC	4258		Db	147057	ACCATCATCCAGGACCTTTTGAACCGTGAAGTATCGCACTGAGCAGCGTCAGCGATGCC	146998	
Db	148137	GGTCCGCAATCACTGTGAGATGTTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC	148078		Qy	5339	ATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTTGGATCGTCCCAACCAAGGGGAGGTG	5398	
Qy	4259	CTGATGCACTCGCGCGTTCGGTCTTCCAGAGTGTGTTGTTTCCCGCAACCAAG	4318		Db	146997	ATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTTGGATCGTCCCAACCAAGGGGAGGTG	146938	
Db	148077	CTGATGCACTCGCGCGTTCGGTCTTCCAGAGTGTGTTGTTTCCCGCAACCAAG	148018		Qy	5399	GTTCACCACTGAGCGGGAAGATCTGTTGGCTTCCCATCTGTCAGCTTTCGCACTTC	5458	
Qy	4319	CGTTTGGTGGCAATGATTTCTTGGCGCGCATTTGGTATGCGATGTTTCCCAAC	4378		Db	146937	GTTCACCACTGAGCGGGAAGATCTGTTGGCTTCCCATCTGTCAGCTTTCGCACTTC	146878	
Db	148017	CGTTTGGTGGCAATGATTTCTTGGCGCGCATTTGGTATGCGATGTTTCCCGGAGC	147958		Qy	5459	CGCACTAAGCTGAGGATGTTTCCCAATGTTGATATCTTTGATGCACTTGGTTTGGACACC	5518	
Qy	4379	CTGTTTACGCTACGAGCTGGCGCCACCATGCGCGCGCAATGCAATGCTGCTC	4438		Db	146877	CGCACTAAGCTGAGGATGTTTCCCAATGTTGATATCTTTGATGCACTTGGTTTGGACACA	146818	
Db	147957	TTGTTGAACGCTACGAGCTGGCGCCACCATGCTCGCGCGCAATGCAATGCTGCTC	147898		Qy	5519	GTAAACCTCAACGGCACGCACTTTAACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA	5578	
Qy	4439	CTGTTTGGTGGATGTTGCTCAAGCTGTTTACAGGGCACCGTGTCTTCTGTGCTGGTG	4498		Db	146817	GTAAACCTCAACGGCACGCACTTTAACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA	146758	
Db	147897	CTGTTTGGTGGATGTTGCTCAAGCTGTTTACAGGGCACCGTGTCTTCTGTGCTGGTG	147838		Qy	5579	GGGGAGCTGCTGTGTGAATTCGATATTTGATGCACTTAAGGCTGCAAGTTATGAGTAACC	5638	
Qy	4499	GTCTCTTGGATTCGGCAAGATCGAGAGTTCCTGCAAGCGACTCATGGCATGCA	4558		Db	146757	GGGGAGCTGCTGTGTGAATTCGATATTTGATGCACTTAAGGCTGCAAGTTATGAGTAACC	146698	
Db	147837	GTCTCTTGGATTCGGCAAGATCGAGAGTTCCTGCAAGCGACTCATGGCATGCA	147778		Qy	5639	ACCGCATGTTGTTTTCGAATTTACAGAAACCGGACCTGTAACACTTTAGCGTTTGGGC	5698	
Qy	4559	GACTTCTGATCAACCCAGTGTGACTCTGCTGCTCACCAGCTTCTTACGTTTCTGCT	4618		Db	146697	ACCGCATGTTGTTTTCGAATTTACAGAAACCGGACCTGTAACACTTTAGCGTTTGGGC	146638	
Db	147777	GACTTCTGATCACTCCAGTGTGACGTTGCTGCTCACCAGGATTCCTTACATTCATCGCC	147718		Qy	5699	GAAATGAAGCGGAGCAACCTGCTCAACGTCGCAAAAGAAAGAGCGGTCGCAACAACA	5758	
Qy	4619	ATTGGTCCAGCAATGCGCTGGTGGTGGTACTTCTGCGCACACAGCTCTGAGGACTCTAT	4678		Db	146637	GAAATGAAGCGGAGCAACCTGCTCAACGTCGCAAAAGAAAGAGCGGTCGCAACAACA	146578	
Db	147717	ATTGGTCCAGCAATGCGCTGGTGGTGGTACTTCTGCGCACACAGCTCTGAGGACTTAT	147658		Qy	5759	CCATAAGTTGAACCTTGAAGTGTTCGCAACACAGGTTAGAGGAGCGTGAAGTCTAGCC	5818	
Qy	4679	GATTTGGTGTGCTGAGTCGCGGCTGCTGTTTTCGGTCTGCTTACTACCAATCGTTATC	4738		Db	146577	CCATAAGTTGAACCTTGAAGTGTTCGCAACACAGGTTAGAGGAGCGTGAAGTCTAGCC	146518	
Db	147657	GATTTGGTGTGCTGAGTCGCGGCTGCTGTTTTCGGTCTGCTTACTACCAATCGTTATC	147598		Qy	5819	ATCTTTGACACCGGTACCCGTCAGCTTGCAGATTTTAAACCTGTTCAACCGGTCATGCC	5878	
Qy	4739	ACTGGTCTGCACCACTGCTCTCCCGCCAAATGAGCTGGAGCTGTTTCAACAGGCTGATCC	4798						

Db 146517 ATCTTTACACCGGTACCGTACGCTTCGAGATTTTAAACCTGTTCAACACGAGTCAAGCC 146458
Qy 5879 TCAGTGTACCTGTGTGGTCCACCCCGCAATCTTCCACCCACATTTGGACATGTCGTTC 5938
Db 146457 TCAGTGTACCTGTGTGGTCCACCCCGCAATCTTCCACCCACATTTGGACATGTCGTTC 146398
Qy 5939 GCAGTACGCTTTGATATTTTGGCCCGCTG 5967
Db 146397 GCAGTACGCTTTGATATTTTGGCCCGCTG 146369

RESULT 3
AX127152/c 349980 bp DNA linear PAT 11-MAY-2001
LOCUS Sequence 7068 from Patent EP1108790.
DEFINITION AX127152 AX114121
ACCESSION AX127152.1 GI:14041140
VERSION
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum
Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE
1 Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
Novel polynucleotides
Patent: EP 1108790-A 7068 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
Location/Qualifiers
1..349980
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
/note="Seq 1 to long (3.309.400) split in 11, seq 7068
2.700.001 3.049.980"
BASE COUNT 81250 a 97718 c 90621 g 80391 t
ORIGIN

Query Match 94.8%; Score 5656.2; DB 6; Length 349980;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

Qy 1 AGTCGTCGACGCGACCACTGATGTGTGTGTCACGAGCTTGCAGGCTTCTACATCT 60
Db 115585 AGTCGTCGACGCGACCACTGATGAGGTGTCACGAGCTTGCAGGCTTCTACATCC 115526

Qy 61 ACCTGCCGTCGCGTGGAGTGGGTGATACGGGTGGGATACGCCGGTGAAGTTCCG 120
Db 115525 ACATGCGCTCGCGTGGAGTGGGTGATACGGGTGGGATACGCCGGTGAAGTTCCG 115466

Qy 121 GAACCCATGCTTCTTGTGGTGGAGGACAGTCCGGTGAGAGTTCCTCAAGTG 180
Db 115465 GAAGCCATGCTTCTTGTGGTGGAGGACAGTCCGGTGAGAGTTCCTCAAGTG 115406

Qy 181 TCTGAGTTTAAAGTTATGATCATCAGCTTGAAGGCTGAGGTAAATTCAGTACCTG 240
Db 115405 TCTGAGTTTAAAGTTATGATCATCAGCTTGAAGGCTGAGGTAAATTCAGTACCTG 115346

Qy 241 CAAACAGAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAGACATCAT 300
Db 115345 CAAACAGAGGCTCAAGTCCGAGATATTAACCTAGATCCGTAGACATAGACATCAT 115286

Qy 301 CGTCTATGCTGCTGGAAGGACCAATACCTCAGAAAGATGCGCAAGTGGTGCAAT 360
Db 115285 CGTCTATGCTGCTGGAAGGACCAATACCTCAGAAAGATGCGCAAGTGGTGCAAT 115226

Qy 361 ATCAAGAAATCAGGTCAAGCAGTTAAAAAATTTAGGGAAGATTTTCCGCCCTCG 420
Db 115225 ATCAAGAAATCAGGTCAAGCAGTTAAAAAATTTAGGGAAGATTTTCCGCCCTCG 115166

Qy 421 GGGTATGATGGCTTCTCCAACTCGAAACGGGATCATCAGCGAATCTCTGGAGAC 480
Db 115165 GGGTATGATGGCTTCTCCAACTCGAAACGGGATCATCAGCGAATCTCTGGAGAC 115106

Qy 481 CAGCACTAAAAACGACGAGGATTCACCCCGAAGTCCCGAGATTTCCCGGTTTATG 540
Db 115105 CAGCACTAAAAACGACGAGGATTCACCCCGAAGTCCCGAGATTTCCCGGTTTATG 115046

Qy 541 ATCTTCATAATCACGGTGGAAACGGTGGCGGTTTCTTACGGGAACGACGAGGCGGA 600
Db 115045 ATCTTCATAATCACGGTGGAAACGGTGGCGGTTTCTTACGGGAACGACGAGGCGGA 114986

Qy 601 GGAACACCGCGCAGTATCACCGCGAAGATGCGACACCGTATGTTGCGAAGCATGGTT 660
Db 114985 GGAATGCGCGCAGTATCACCGCGAAGATGCGACACCGTATGTTGCGAAGCATGGTT 114926

Qy 661 CGCGCCCGGCTGACGCATTCGCGAGCGAGGTGGAAACCTTATTCCTTGTGTGAAGAG 720
Db 114925 CGCGCCCGGCTGACGCATTCGCGAGCGAGGTGGAAACCTTATTCCTTGTGTGAAGAG 114866

Qy 721 TCTGCTGTGGCGGATTCACCTCGAGGGCCCTTTCATCAACGATGCGGTTGTGTGCTC 780
Db 114865 GCTGCTGTGGCGGATTCACCTCGAGGGTTCCTTTCATCAACGATGCGGTTGTGTGCTC 114806

Qy 781 AAAACCCGGATTTCAATTTTCCCGGCAACCAACAGATCTTCCCGGGTGTATCCATCGG 840
Db 114805 AAAACCCGGATTTTATTTTCCCGCAACCAACAGATCTTCCCGAGGTATCCATCGG 114746

Qy 841 GAAAGGTTGGATCAAAATCGATGACAGTACGCGCCGGAACTCACAATCTTTCTGAGCTC 900
Db 114745 GAAAGGTTGGATCAAAATCGATGACAGTACGCGCCGGAACTCACAATCTTTCTGAGCTC 114686

Qy 901 TCGATCTCTGCGAGCGCACCATCATTCATTCCTTCTCGGGCACACTGATGACAGATTG 960
Db 114685 TCGATCTCTGCGAGCGCACCATCATTCATTCCTTCTCGGGCACACTGATGACAGATTG 114626

Qy 961 ATACCACTACGAGCGCAATTTGCTGGCTTAAAGAGAAATGTGACGGTACGCGTACGC 1020
Db 114625 ATACCACTACGAGCGCAATTTGCTGGCTTAAAGAGAAATGTGACGGTACGCGTACGC 114566

Qy 1021 ATTTGTTCAATGCGATGCTCGCTGCATCATAGGGCTCCCGGAGCGTGGCGCTTGC 1080
Db 114565 ATTTGTTCAATGCGATGCTCGCTGCATCATAGGGATCCCGGAGCGTGGCGCTTGC 114506

Qy 1081 TTCTCGCGACGCTGCGGGGAGCGCATATGTTGAGTTGATTCGCGAGCGGTGCATTTGG 1140
Db 114505 TTCTCGCGACGCTGCGGGGAGCGCATATGTTGAGTTGATTCGCGAGCGGTGCATTTGG 114446

Qy 1141 CCATGGAAGCGTGCATCTAGCTCGTTCCAAACACGCTTTTTCATCAGCGAGCGCATGG 1200
Db 114445 CCATGGAAGCGTGCATCTAGCTCGTTCCAAACACGCTTTTTCATCAGCGAGCGCATGG 114386

Qy 1201 AAGCCGCGGAATGCCAGACGGTGAATATTTGGGCGTTTGAACGTCACCGTCAACCG 1260
Db 114385 AAGCCGCGGAATGCCAGACGGTGAATATTTGGGCGTTTGAACGTCACCGTCAACCG 114326

Qy 1261 ATGGAGTGGCGGCTGCTCGCGATGGCGGCCATCCCGGGGACACGACACTAGCGA 1320
Db 114325 ATGGCGTGGCGGCTGCTCGCGATGGCGGCCATCCCGGGGATCCAGCAGACTAGCGA 114266

Qy 1321 GTCAATGCTGTCACCGTGCAGAGGGTATGAGCTTATTCGACGGGACCGCTCCACCT 1380
Db 114265 GTCAATGCTGTCACCGTGCAGAGGGTATGAGCTTATTCGACGGGACCGCTCCACCT 114206

Qy 1381 CAACCGTGGCGGTAAATTTCTCGGCTTGGGATCAGGAAATCGCTAAATCCAAACCTG 1440
Db 114205 CAACCGTGGCGGTAAATTTCTCGGCTTGGGATCAGGAAATCGCTAAATCCAAACCTG 114146

Qy 1441 CAAATTTTGTGCTTTGACTCAACGCGGAGGTGCAAAAGGTCCATTTAGGTCAATCAAG 1500
Db 114145 TAAATTTTGTGCTTTGACTCAACGCGGAGGTGCAAAAGGTCCATTTAGACCAATCAAG 114086

Qy 1501 TACTTTAAGTACGAGTAAAAATCTCTGAT -TTTAAAGAGTCCCACTGGAATCACT 1559
Db 114085 TAAATTTAAGTACGAGTAAAAATCTCTGAT -TTTAAAGAGTCCCACTGGAATCACT 114026

Db	111866	GCCTTTGAACAACATTCATGCTGATATATTTTGGCTCTCTCCGGTTAAGGAAATTC	111807
Qy	3779	ATGGACCATTAAGACCTCGCGCAACGCATCTCGCGACATTTGGCGCGCAAGACATTC	3838
Db	111806	ATGGACCATTAAGACCTCGCGCAACGCATCTCGCGACATTTGGCGCGCAAGACATTC	111747
Qy	3839	GTGCGCGCGCACACTGTGCAACCGCTTACGCGCTCGTGTCTCAAGACACCAAGGATG	3898
Db	111746	GTGCGCGCGCACACTGTGCAACCGCTTACGCGCTCGTGTCTCAAGACACCAAGGATG	111687
Qy	3999	GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAAGGCACCTTTGAAACTGCGCGCATG	3958
Db	111686	GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAAGGCACCTTTGAAACTGCGCGCATG	111627
Qy	3959	TTCCAGATCATCTGCGCGCAGCGGATGCGATCATGTTTCAAGAACTCGATGACGA	4018
Db	111626	TTCCAGATCATCTGCGCGCAGCGGATGCGATCATGTTTCAAGAACTCGATGACGA	111567
Qy	4019	ACCTTCCAAAGACATCGCTGTGTCACAGACAGCTCAAGAGATTTGTGGCTAAACACGCC	4078
Db	111566	ACCTTCCAAAGACATCGCTGTGTCACAGACAGCTCAAGAGATTTGTGGCTAAACACGCC	111507
Qy	4079	AACCTGGTTCAGCGCTGCTGTAAGGATTTGGCGGACATTTTCGTCGCGCTGATTCGAATC	4138
Db	111506	AACCTGGTTCAGCGCTGCTGTAAGGATTTGGCGGACATTTTCGTCGCGCTGATTCGAATC	111447
Qy	4139	TTGGTTGGTGGCGCTGCTGTCATGCTGATCAACAATGTGTTGGTGGCGAGGATCTGTTTC	4198
Db	111446	TTGGTTGGTGGCGCTGCTGTCATGCTGATCAACAATGTGTTGGTGGCGAGGATCTGTTTC	111387
Qy	4199	GGTCCGCAACTGCTGTTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC	4258
Db	111386	GGTCCGCAACTGCTGTTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC	111327
Qy	4259	CTGATGTCATCTGCGCGCTGCGGTGCTTCCAGTGTGTTGGTGGTTCACCGCAACCAAG	4318
Db	111326	CTGATGTCATCTGCGCGCTGCGGTGCTTCCAGTGTGTTGGTGGTTCACCGCAACCAAG	111267
Qy	4319	CGTTTCGGTGGCAATGAGTTCCTGGCGCGCGGATGTTGATGCGGATGTTGTTCCCAACC	4378
Db	111266	CGTTTCGGTGGCAATGAGTTCCTGGCGCGCGGATGTTGATGCGGATGTTGTTCCCGGAGC	111207
Qy	4379	CTGGTTAACGGCTACGACGTGGCGCCACCACATGACCGCGCGGAAATGCCAATGTGGTCC	4438
Db	111206	TTGGTGAACGGCTACGACGTGGCGCCACCACATGCTGCGCGCGGAAATGCCAATGTGGTCC	111147
Qy	4439	CTGTTTGGTGGATGTTGCTCAAGCTGTTTACAGGGCACCGTGTCTGCTGGTGGT	4498
Db	111146	CTGTTTGGTGGATGTTGCTCAAGCTGTTTACAGGGCACCGTGTCTGCTGGTGGT	111087
Qy	4499	GTCTCTTGGATTTCTGCAACGATCGAGAGTTCCTGCACAGGACCTCATGGGCATGCA	4558
Db	111086	GTCTCTTGGATTTCTGCAACGATCGAGAGTTCCTGCACAGGACCTCATGGGCATGCA	111027
Qy	4559	GACTTCTGTATCACCCAGTGTGTGACFTGCTGCTCACCGGCTTCCTTACGTTTCATGTCT	4618
Db	111026	GACTTCTGTATCACCCAGTGTGTGACFTGCTGCTCACCGGCTTCCTTACGTTTCATGTCT	110967
Qy	4619	ATTGGTCCAGCAATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	4678
Db	110966	ATTGGTCCAGCAATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	110907
Qy	4679	GATTTCGGTGGTCCAGTCGGCGGTGCTGCTTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT	4738
Db	110906	GATTTCGGTGGTCCAGTCGGCGGTGCTGCTTCTCGGTGCTGCTTCTACTCAACCAATCGTCATC	110847
Qy	4739	ACTGGTCTGCACAGTCTCTCCGCCAATTTGAGCTGGAGCTGTTCAACAGGGTGGATCC	4798
Db	110846	ACTGGTCTGCACAGTCTCTCCGCCAATTTGAGCTGGAGCTGTTTAAACAGGGTGGATCC	110787
Qy	4799	TTTCTCTGCAACCGCATCCATGGCCAAATATCGCGAGGGTGCAGCATGTTGGGAGTG	4858
Db	110786	TTTCTCTGCAACCGCATCTATGGCTAATATCGCCCGAGGGTGGGCAATGTTGGCAGTG	110727
Qy	4859	TTCTTCTCCTAGGCAAGAGTGAAAAGCTCAAGGCGCTTCAGGTGCTTCAGGTGCTCCTCGCT	4918
Db	110726	TTCTTCTCCTAGGCAAGAGTGAAAAGCTCAAGGCGCTTCAGGTGCTTCAGGTGCTCCTCGCT	110667
Qy	4919	GTTCTTGGTATTACAGAGCTCGCATCTTCGGTGTGAACCTTCGGCTCGCTCGCTGGCGGTC	4978
Db	110666	GTTCTTGGTATTACAGAGCTCGCATCTTCGGTGTGAACCTTCGGCTCGCTCGCTGGCGGTC	110607
Qy	4979	TACATTTGGTATCGGTACCGCAGCTATCGGTGGCGGCTTTGATGTCATCTTTGATATCAAG	5038
Db	110606	TACATTTGGTATCGGTACCGCAGCTATCGGTGGCGGCTTTGATGTCATCTTTTAAATCAAG	110547
Qy	5039	GCAGTTGGCTTGGCGGCTGCGAGTTTCTTGGTGTGTTGTTTATTGATGCTCCAGATATG	5098
Db	110546	GCAGTTGGCTTGGCGGCTGCGAGTTTCTTGGTGTGTTGTTTATTGATGCTCCAGATATG	110487
Qy	5099	GTCTATGTTCTTGGTGGCGGCTAGTTACCTTTGTCATCGGATTCGGCGCAGCATTCCT	5158
Db	110486	GTCTATGTTCTTGGTGGCGGCTAGTTACCTTTGTCATCGGATTCGGCGCAGCATTCCT	110427
Qy	5159	TATGGCTTTACTTGGTTCGCGCAGCGCAGCATTCAGATGCAACGCTGCTCTCA	5218
Db	110426	TATGGCTTTACTTGGTTCGCGCAGCGCAGCATTCAGATGCAACGCTGCTCTCA	110367
Qy	5219	GTGCTGTCAGGAACACCAAGCCGAGCAGAACCCCGCAGAAATTTTCAAAAGATTC	5278
Db	110366	GTGCTGTCAGGAACACCAAGCCGAGCAGAACCCCGCAGAAATTTTCAAAAGATTC	110307
Qy	5279	ACCATCATCCAGCAGCTTTGACCGGTGAAGCTATCGCATGAGCAGGCTGAGCGATGCC	5338
Db	110306	ACCATCATCCAGCAGCTTTGACCGGTGAAGCTATCGCATGAGCAGGCTGAGCGATGCC	110247
Qy	5339	ATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTGGGATCGTCCCGCAGGAGGGGAGTG	5398
Db	110246	ATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTGGGATCGTCCCGCAGGAGGGGAGTG	110187
Qy	5399	GTTTCCACAGTCAGCGGAAAGATCGTGGTGGCGCTTCCCATCTGGTTCAGCGCTTCGACGC	5458
Db	110186	GTTTCCACAGTCAGCGGAAAGATCGTGGTGGCGCTTCCCATCTGGTTCAGCGCTTCGACGC	110127
Qy	5459	CGCATTAAGGCTGAGGATGGTTCCCATGTGGGATATCTTGTGATGACATTTGGTTCGACACC	5518
Db	110126	CGCATTAAGGCTGAGGATGGTTCCCATGTGGGATATCTTGTGATGACATTTGGTTCGACACC	110067
Qy	5519	GTAACCTCAACGGCAGCGACCTTTACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA	5578
Db	110066	GTAACCTCAACGGCAGCGACCTTTACCGCTGGAAGAGCAGGCGGATGAAGTCAAGCA	110007
Qy	5579	GGGAGCTGCTGTGTAATTCGATATTTGATGCCATTAAGGCTGAGGTTATGAGTAAACC	5638
Db	110006	GGGAGCTGCTGTGTAATTCGATATTTGATGCCATTAAGGCTGAGGTTATGAGTAAACC	109947
Qy	5639	ACGCGATTTGTTTTCGAATTAAGAAACCGGACCTGTAAACACTTACGTTTGGG	5698
Db	109946	ACGCGATTTGTTTTCGAATTAAGAAACCGGACCTGTAAACACTTACGTTTGGG	109887
Qy	5699	GAAATTTGAAGCGGGAGCCAACTGCTCAACGTGCAAGAAAGAAAGAGCGGTGCCAACA	5758
Db	109886	GAAATTTGAAGCGGGAGCCAACTGCTCAACGTGCAAGAAAGAAAGAGCGGTGCCAACA	109827
Qy	5759	CGATAAGTTGAACCTTTGAGTGTTCGACACAGGTTAGACTAGGGGAGTGACTCTAGC	5818
Db	109826	CGATAAGTTGAACCTTTGAGTGTTCGACACAGGTTAGACTAGGGGAGTGACTCTAGC	109767
Qy	5819	ATCTTTGACACCGGTACCGGTACGTTTCGAGATTTTAAACCTTTTCAACAGGTCATGCC	5878
Db	109766	ATCTTTGACACCGGTACCGGTACGTTTCGAGATTTTAAACCTTTTCAACAGGTCATGCC	109707
Qy	5879	TGCGTGTACCTGTGTGGTCCCGCCGCAATCTTCAACCCCATGTTGGACATGTTCTGTTCA	5938
Db	109706	TGCGTGTACCTGTGTGGTCCCGCCGCAATCTTCAACCCCATGTTGGACATGTTCTGTTCA	109647

QY	5939	GCAGTAGCGTTTGATATTGTCGGCCGCTG	5967
Dd	109646	GCAGTAGCGTTTGATATTGTCGGCCGCTG	109618
RESULT 4			
E11760			
LOCUS	E11760	6911 bp	DNA linear PAT 29-SEP-1997
DEFINITION	Base sequence of sucrase gene.		
ACCESSION	E11760		
VERSION	E11760.1	GI:22025396	
KEYWORDS	JP 1996196280-A/1.		
SOURCE	Corynebacterium glutamicum		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.		
REFERENCE	1 (bases 1 to 6911)		
AUTHORS	Yoshihara,M., Otona,K., Nagase,K., Tsuchiya,M., Matsui,Y., Yoshihara,Y. and Nakamatsu,W.		
TITLE	SUCROSE GENE ORIGINATED FROM CORYNEFORM BACTERIUM AND ITS		
JOURNAL	Patent: JP 1996196280-A 1 06-AUG-1996;		
COMMENT	AJINOMOTO CO INC		
OS	Brevibacterium lactofermentum		
PN	JP 1996196280-A/1		
PD	06-AUG-1996		
PF	30-JAN-1995	JP 1995012361	
PI	SUGIMOTO MASAKAZU, OTONA KIKOKO, NAGASE KAZUO, TSUCHIYA MAKOTO, PI MATSUI YUTAKA, YOSHIIHARA YASUHIKO, NAKAMATSU WATARU PC		
C12N15/09,C12N1/21,C12P13/06/C12N5/24,(C12N1/21,C12R1:15); CC			
strandedness:	Double;		
CC topology:	Linear;		
FH Key	Location/Qualifiers		
FT source	1..6911		
FT	/organism='Brevibacterium lactofermentum' FT		
FT	/strain='ATCC13869'		
FT	/plasmid='PSSM30BS'		
FT CDS	342..1505		
FT	/product='A protein having high homology to FT		
FT	N-acetylglucosamine-6-phosphate deacetylase', FT		
CDS	2338..3609		
FT	/product='Sucrase',		
FT	4438..5358		
FT	/product='A protein high homology to FT		
FT	UDP-N-acetylmuramoyl		
FT	alanyl-D-glutamyl-meso-6-diaminopimelate FT		
FT	'synthetase',		
FT CDS	5570..6577		
FT	/product='A protein having hmology to FT		
FT	phospho-N-acetylmuramoyl pentapeptide FT		
transferase',			
source	Location/Qualifiers		
i..6911			
/organism='Corynebacterium glutamicum'			
/mol_type='genomic DNA'			
/db_xref='taxon:1718'			
BASE COUNT	1575 a 1784 c 1892 g 1660 t		
ORIGIN			
Query Match	61.8%; Score 3687; DB 6; Length 6911;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 3687; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	AGTCGTGACGCCACCATTGATGGTGGTCCAGCGCTGCGAGGCTTCTACATCT	60
Dd	1	AGTCGTGACGCCACCATTGATGGTGGTCCAGCGCTTCTGAGGCTTCTACATCT	60
QY	61	ACGCTCCCGTCGGCTGGAGTGGGTCAATTACGGGTGGGATCACCGCGTGAAGAATTGCG	120
Dd	61	ACGCTCCCGTCGGCTGGAGTGGGTCAATTACGGGTGGGATCACCGCGTGAAGAATTGCG	120
QY	121	GAAACCATGGTCTCCTTGTGGTGTGAGGAAACAGTGC GG GTGAGAAGTTTTTCAAGT	180

QY	1261	ATGAGTCCCGCTGTGCGGATGSGCGCGCCATCGCCGGGGGACCAGCACACTAGCGA	1320
Db	1261	ATGAGTCCCGCTGTGCGGATGSGCGCGCCATCGCCGGGGGACCAGCACACTAGCGA	1320
QY	1321	GTCACTGTGTGACACAGTGGCGAGGTATGAGCGCTATATGACGCGACCCCTCCACACT	1380
Db	1321	GTCACTGTGTGACACAGTGGCGAGGTATGAGCGCTATATGACGCGACCCCTCCACACT	1380
QY	1381	CAACGTCGCCCTAAAATTCCTGCTTTCGGCTGTGGCGATCAGAAATCGTAAATCCAAACCTG	1440
Db	1381	CAACGTCGCCCTAAAATTCCTGCTTTCGGCGATCAGAAATCGTAAATCCAAACCTG	1440
QY	1441	CAAAATTTGTGCTTTGACTCAACAGCGCGAGGTGCAAAAGGTCCATTTAGGTATCAACG	1500
Db	1441	CAAAATTTGTGCTTTGACTCAACAGCGCGAGGTGCAAAAGGTCCATTTAGGTATCAACG	1500
QY	1501	TACTTTAAGTAGAGTAAACATTCCTGATTTTAAAGGAGTCCACCATGGAATCACTA	1560
Db	1501	TACTTTAAGTAGAGTAAACATTCCTGATTTTAAAGGAGTCCACCATGGAATCACTA	1560
QY	1561	TCGCAAGACGAGCAAGAGTCGGCAAGAGTGTGAGTCTTAATGCAACCTTCGCCA	1620
Db	1561	TCGCAAGACGAGCAAGAGTCGGCAAGAGTGTGAGTCTTAATGCAACCTTCGCCA	1620
QY	1621	ACAAGGTGGAACCTTGGGCTTGAACAGAGTCTCACACTGAGTACCTAACAGAGC	1680
Db	1621	ACAAGGTGGAACCTTGGGCTTGAACAGAGTCTCACACTGAGTACCTAACAGAGC	1680
QY	1681	TCAATCGCATGTATGAAGTGGGAGTGTTCATTCAGAACTGCAAGGCATTTGTTGG	1740
Db	1681	TCAATCGCATGTATGAAGTGGGAGTGTTCATTCAGAACTGCAAGGCATTTGTTGG	1740
QY	1741	ATGAATACCTGGACCTAACCCGTGAGTGAAGAGTCTTAAACAGTCTTAAACAGTTCGCAAG	1800
Db	1741	ATGAATACCTGGACCTAACCCGTGAGTGAAGAGTCTTAAACAGTCTTAAACAGTTCGCAAG	1800
QY	1801	AGTCACTGACACATCGACATCGTGTATGAAGAGTCTACAGCCAGATGGTGCAAACT	1860
Db	1801	AGTCACTGACACATCGACATCGTGTATGAAGAGTCTACAGCCAGATGGTGCAAACT	1860
QY	1861	CTGATCCATACGAGAGTCCAGAGTATGAGGCAAGATCGCTGCAGAAATCCGTTGAAG	1920
Db	1861	CTGATCCATACGAGAGTCCAGAGTATGAGGCAAGATCGCTGCAGAAATCCGTTGAAG	1920
QY	1921	TTCAAAATCCTTGGCATCGCGGAAACGGCACATCGCTTTTCATTGAACCATCATCTCTCT	1980
Db	1921	TTCAAAATCCTTGGCATCGCGGAAACGGCACATCGCTTTTCATTGAACCATCATCTCTCT	1980
QY	1981	GTCAAGACTGCAAAAGTCCAGGCGTGCACCTTAAACCTGTGGAGGACACGCTCGATT	2040
Db	1981	GTCAAGACTGCAAAAGTCCAGGCGTGCACCTTAAACCTGTGGAGGACACGCTCGATT	2040
QY	2041	CTTCAACACATCGAAGAGTCCCAACCCAGCGCTCACAGGTTTGGGCATTTGTC	2100
Db	2041	CTTCAACACATCGAAGAGTCCCAACCCAGCGCTCACAGGTTTGGGCATTTGTC	2100
QY	2101	CCGCGGCAAAACATCGTGTGGTGGCACTGGTGAAGGAAAGCCGACGCTCCCGGG	2160
Db	2101	CCGCGGCAAAACATCGTGTGGTGGCACTGGTGAAGGAAAGCCGACGCTCCCGGG	2160
QY	2161	AACGTGGAGGCCAGTACCTGCTTCTGTCGCGAGTTCATCTCTGTAGATGCACACAT	2220
Db	2161	AACGTGGAGGCCAGTACCTGCTTCTGTCGCGAGTTCATCTCTGTAGATGCACACAT	2220
QY	2221	GCCACCATCATGTTGGATGAAGCAGCAGTATCCAAAGTGGAAACCGTGTACCTACCG	2280
Db	2221	GCCACCATCATGTTGGATGAAGCAGCAGTATCCAAAGTGGAAACCGTGTACCTACCG	2280
QY	2281	TCTCATGGAGCAATTTAAAGCTGGCTAGAAACAAAGAAAGTACTGTGTGGGCTATG	2340
Db	2281	TCTCATGGAGCAATTTAAAGCTGGCTAGAAACAAAGAAAGTACTGTGTGGGCTATG	2340
QY	2341	CACACAGAACTTTCAGTTTGGGCTTCGCTACCATGTGACTCTCCGACAGGCGAGCTC	2400
Db	2341	CACACAGAACTTTCAGTTTGGGCTTCGCTACCATGTGACTCTCCGACAGGCGAGCTC	2400
QY	2401	AATGATCCCAACGGAATGTAGTGTGATGAGATACCTTCCACGCTCTACTACACAGCAT	2460
Db	2401	AATGATCCCAACGGAATGTAGTGTGATGAGATACCTTCCACGCTCTACTACACAGCAT	2460
QY	2461	CCAGTTTTCCTTCGCAACCAAGCGCGCTGGGCTCACACCAACCGCGTTCGAC	2520
Db	2461	CCAGTTTTCCTTCGCAACCAAGCGCGCTGGGCTCACACCAACCGCGTTCGAC	2520
QY	2521	GGACGCGCAGTATTCAGTGGACGACCTGCGCGACGCTCTTACCGGATGATCTCTAT	2580
Db	2521	GGACGCGCAGTATTCAGTGGACGACCTGCGCGACGCTCTTACCGGATGATCTCTAT	2580
QY	2581	GACCTGGATGTATTCGCTGGAGCGCTATTTACTAGCGACACACTTAACTTTTC	2640
Db	2581	GACCTGGATGTATTCGCTGGAGCGCTATTTACTAGCGACACACTTAACTTTTC	2640
QY	2641	TACACGGCAACCTTAAATTTGACGGAAGCGCGGCCACCCCAAAACCTTGTGCAAGTC	2700
Db	2641	TACACGGCAACCTTAAATTTGACGGAAGCGCGGCCACCCCAAAACCTTGTGCAAGTC	2700
QY	2701	GAGGACCAACTTGGCTGATGGCGGCTTTCATGCGCTTCGCTTAAATTCGCTTATC	2760
Db	2701	GAGGACCAACTTGGCTGATGGCGGCTTTCATGCGCTTCGCTTAAATTCGCTTATC	2760
QY	2761	GACGACCCGCGAGCGGTTTACACCCCATTTACCGCATCCCATGATCAGCCCTGATGGT	2820
Db	2761	GACGACCCGCGAGCGGTTTACACCCCATTTACCGCATCCCATGATCAGCCCTGATGGT	2820
QY	2821	GATGTTGAAACATGGTCTTGGGCGCCACGCAAAACCTCACCGGTGACGCGTCTA	2880
Db	2821	GATGTTGAAACATGGTCTTGGGCGCCACGCAAAACCTCACCGGTGACGCGTCTA	2880
QY	2881	TACCGCTCACAGATCTTGAACCTTGGAAATTTCCCGTGAATCACTTTGACCTCAGT	2940
Db	2881	TACCGCTCACAGATCTTGAACCTTGGAAATTTCCCGTGAATCACTTTGACCTCAGT	2940
QY	2941	GATGACAACTTGTCTGCTCTGATCTGCTTCCGATGCTACATGTGGGAATGCCCC	3000
Db	2941	GATGACAACTTGTGTCTGCTCTGATCTGCTTCCGATGCTACATGTGGGAATGCCCC	3000
QY	3001	AACCTTTTACGCTTCGGGATGAAGAACTGGCGAAGATCTCGACGCTGCTATTTCTGT	3060
Db	3001	AACCTTTTACGCTTCGGGATGAAGAACTGGCGAAGATCTCGACGCTGCTATTTCTGT	3060
QY	3061	CCACAGGATTTGGACCGAATCCACGATGAGTTACTACTACGCAAGCTTGACCACTGC	3120
Db	3061	CCACAGGATTTGGACCGAATCCACGATGAGTTACTACTACGCAAGCTTGACCACTGC	3120
QY	3121	GGATATGCTGTGCAAGGCTTGAAGGAAACGCTTCCCGCTTCTGCGAGGATTCAGGAG	3180
Db	3121	GGATATGCTGTGCAAGGCTTGAAGGAAACGCTTCCCGCTTCTGCGAGGATTCAGGAG	3180
QY	3181	CTGGATTTCCGCCATGAATTTACGACCGCGAGGTTGCAGTAAACGGTTCTGTAGCCTGG	3240
Db	3181	CTGGATTTCCGCCATGAATTTACGACCGCGAGGTTGCAGTAAACGGTTCTGTAGCCTGG	3240
QY	3241	CTCGTGGGCTGGATGGGCTGCCCGCAGGATGATACCCAAACAGTTGCAACAGGAAGA	3300
Db	3241	CTCGTGGGCTGGATGGGCTGCCCGCAGGATGATACCCAAACAGTTGCAACAGGAAGA	3300
QY	3301	TGGGTGCATCGCTGACTGTGCCCGCGAAGCTTCATTTGCGCAACCCACGCGATCTACAA	3360
Db	3301	TGGGTGCATCGCTGACTGTGCCCGCGAAGCTTCATTTGCGCAACCCACGCGATCTACAA	3360
QY	3361	GACCTCCTTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGTATAGGTTCTGAACCT	3420
Db	3361	GACCTCCTTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGTATAGGTTCTGAACCT	3420
QY	3421	GTCCGAGTAGACATCCGAGGCAATTTCCCTCCGAGTGGGATGGTGTCTGTCTGTG	3480

Db 1501 TACTTTAAGTACGAGTAAACTATCCTGATTTTAAAGGAGTCCCAACCATGGAATCACTA 1560
Qy 1561 TCTGCAAGACGAGCAAGAGTGGGCAAGACAGTGTGAGTCCCTAAATCGCAACCCCTTGGCCA 1620
Db 1561 TCTGCAAGACGAGCAAGAGTGGGCAAGACAGTGTGAGTCCCTAAATCGCAACCCCTTGGCCA 1620
Qy 1621 ACAAAGGTGGAACCTTGGGCTTGGCAACAGGATCCCTCAACACATGAGTACCTACCAAGAGC 1680
Db 1621 ACAAAGGTGGAACCTTGGGCTTGGCAACAGGATCCCTCAACACATGAGTACCTACCAAGAGC 1680
Qy 1681 TCATTCCCATGTATGAGCTTGGGAGTGTCAATTCAGAAGTCTCAAGAGGCAATCTTGTGTTGG 1740
Db 1681 TCATTCCCATGTATGAGCTTGGGAGTGTCAATTCAGAAGTCTCAAGAGGCAATCTTGTGTTGG 1740
Qy 1741 ATGAATACGTGGGACTAACCCGTGACGATGAAGAAACAGCTACTTTAAACCATTCGCAAG 1800
Db 1741 ATGAATACGTGGGACTAACCCGTGACGATGAAGAAACAGCTACTTTAAACCATTCGCAAG 1800
Qy 1801 AGTTCACTGACCAACATCGACATCGTTGATGAAGAGTCTTACAGCCAGATGGTGCAAAAC 1860
Db 1801 AGTTCACTGACCAACATCGACATCGTTGATGAAGAGTCTTACAGCCAGATGGTGCAAAAC 1860
Qy 1861 CTGATCCATAGCAAGACGTCAGAGTATGAGCAAGATCGCTGCAGATCCGTTGAAG 1920
Db 1861 CTGATCCATAGCAAGACGTCAGAGTATGAGCAAGATCGCTGCAGATCCGTTGAAG 1920
Qy 1921 TTCAAACTCTTGGCATCGGCGGAAAGGACATCGCTTTCATTGAACCATCATCTTCTCT 1980
Db 1921 TTCAAACTCTTGGCATCGGCGGAAAGGACATCGCTTTCATTGAACCATCATCTTCTCT 1980
Qy 1981 GTCAAGGACTGACAAGGTCAGAGGCTGCACCTTAAACCTGTGGAGCAACACCTCGATT 2040
Db 1981 GTCAAGGACTGACAAGGTCAGAGGCTGCACCTTAAACCTGTGGAGCAACACCTCGATT 2040
Qy 2041 CTTCAACACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGGCATTTGTC 2100
Db 2041 CTTCAACACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGGCATTTGTC 2100
Qy 2101 CCGCGCGCAAAACATCGTGTGGCACTGTGGAAGAAAGCGACGCCATCCGCGG 2160
Db 2101 CCGCGCGCAAAACATCGTGTGGCACTGTGGAAGAAAGCGACGCCATCCGCGG 2160
Qy 2161 AACTGTGGAAGGCCAGTACTGCTTCTTCCAGGTTCCATCCTGTAGATGCACAACAT 2220
Db 2161 AACTGTGGAAGGCCAGTACTGCTTCTTCCAGGTTCCATCCTGTAGATGCACAACAT 2220
Qy 2221 GCCACCATCATCTGTGATGAGCAGCAGTATCCAGCTGGAAGACGCTGATCACTACCG 2280
Db 2221 GCCACCATCATCTGTGATGAGCAGCAGTATCCAGCTGGAAGACGCTGATCACTACCG 2280
Qy 2281 TCTCATGAGCAATTAAGCTGGCTAGAAACAAAAGGAAAGTACTGTGTGGGGCTATG 2340
Db 2281 TCTCATGAGCAATTAAGCTGGCTAGAAACAAAAGGAAAGTACTGTGTGGGGCTATG 2340
Qy 2341 CACACAGACTTTCACGTTTGGCCCTCGCTACCATGTGACTCTCCGAGGCGAGGCTC 2400
Db 2341 CACACAGACTTTCACGTTTGGCCCTCGCTACCATGTGACTCTCCGAGGCGAGGCTC 2400
Qy 2401 AATGATCCCAACGGAATGTACGTGATGAGATACCCCTCCAGCTACTACTACCAAGCAGAT 2460
Db 2401 AATGATCCCAACGGAATGTACGTGATGAGATACCCCTCCAGCTACTACTACCAAGCAGAT 2460
Qy 2461 CCAGGTTTCCCTTCGCAACAAAGCGACCGGTGGGCTCACACACCGCCGCTTGACC 2520
Db 2461 CCAGGTTTCCCTTCGCAACAAAGCGACCGGTGGGCTCACACACCGCCGCTTGACC 2520
Qy 2521 GGACCGCAGGATTCAGTGGACGACCTCGCCAGCCTCTTACCCGGATGATCCTAT 2580
Db 2521 GGACCGCAGGATTCAGTGGACGACCTCGCCAGCCTCTTACCCGGATGATCCTAT 2580
Qy 2581 GACCTGATGATGCTATTCGGTGGAGCGGATTTTACTACGGCACACTTAAACCTTTC 2640
Db 2581 GACCTGATGATGCTATTCGGTGGAGCGGATTTTACTACGGCACACTTAAACCTTTC 2640

Qy 2641 TACACGGCAACCTTAAATTTACGGAAGAGCCCGCCACCCAAACCTTGTGCAAGTC 2700
Db 2641 TACACGGCAACCTTAAATTTACGGAAGAGCCCGCCACCCAAACCTTGTGCAAGTC 2700
Qy 2701 GAGGACCCAACTGGGCTGATGGCGGCATTTATCGCCCTTCCCTTAAATTCGGTTATC 2760
Db 2701 GAGGACCCAACTGGGCTGATGGCGGCATTTATCGCCCTTCCCTTAAATTCGGTTATC 2760
Qy 2761 GAGGACCCCGCAGCGGTTTACACCCCATTTACCGGATCCCATGATCAGCCCTGATGGT 2820
Db 2761 GAGGACCCCGCAGCGGTTTACACCCCATTTACCGGATCCCATGATCAGCCCTGATGGT 2820
Qy 2821 GATGTTGGAACATGGTTCTTGGGCGCAACCGGAAACCTCACCGGTGCAGCGTTCTA 2880
Db 2821 GATGTTGGAACATGGTTCTTGGGCGCAACCGGAAACCTCACCGGTGCAGCGTTCTA 2880
Qy 2881 TACCGTCCACAGATCTTGAACCTGGGAATTTCCCGGTGAATCACCTTTGACCTCAGT 2940
Db 2881 TACCGTCCACAGATCTTGAACCTGGGAATTTCCCGGTGAATCACCTTTGACCTCAGT 2940
Qy 2941 GATGCACAACCTGGTTCTCTCTGATCTCGTTCCTCGATGGGAAATGCCCC 3000
Db 2941 GATGCACAACCTGGTTCTCTCTGATCTCGTTCCTCGATGGGAAATGCCCC 3000
Qy 3001 AACCTTTTACGCTTCGCGATGAAGAACTGGCGAAGATCTCGAGCTGCTGATTCTGT 3060
Db 3001 AACCTTTTACGCTTCGCGATGAAGAACTGGCGAAGATCTCGAGCTGCTGATTCTGT 3060
Qy 3061 CCACAGGATTTGACCGGAATCCACGATGAGTACTCTACTAGCAAGCTCTGACCAAGTC 3120
Db 3061 CCACAGGATTTGACCGGAATCCACGATGAGTACTCTACTAGCAAGCTCTGACCAAGTC 3120
Qy 3121 GGATATGCTCGCAACAGCTTGAAGAGCAGCCTTCCCGCTTCCGAGGATTCAGCGAG 3180
Db 3121 GGATATGCTCGCAACAGCTTGAAGAGCAGCCTTCCCGCTTCCGAGGATTCAGCGAG 3180
Qy 3181 CTGGATTTCCGCGCATGAATTTCTACGACCGCAGGTTGCAGTAAACGGTTCTGATCCCTGG 3240
Db 3181 CTGGATTTCCGCGCATGAATTTCTACGACCGCAGGTTGCAGTAAACGGTTCTGATCCCTGG 3240
Qy 3241 CTCGTGGGCTGATGGGGTCCCGCGCAGGATGATCACCCAAACAGTTGCAAGAGGA 3300
Db 3241 CTCGTGGGCTGATGGGGTCCCGCGCAGGATGATCACCCAAACAGTTGCAAGAGGA 3300
Qy 3301 TGGGTGCACTGCTGACTGTGCGCCGCAAGCTTCAATTTCCGCAACCGGATCTACCAA 3360
Db 3301 TGGGTGCACTGCTGACTGTGCGCCGCAAGCTTCAATTTCCGCAACCGGATCTACCAA 3360
Qy 3361 GAGCTCCTTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGTATTAGGTTCTGAACCT 3420
Db 3361 GAGCTCCTTCTCCAGAGGGGAGTCCGGGGTAAATCAGATCTGTATTAGGTTCTGAACCT 3420
Qy 3421 GTCCGAGTAGACATCCGAGGCAATTTTCCCTCGAGTGGGATGCTCGGTTGTCTGTG 3480
Db 3421 GTCCGAGTAGACATCCGAGGCAATTTTCCCTCGAGTGGGATGCTCGGTTGTCTGTG 3480
Qy 3481 GATCCTGATGTTGATCGTGCCTAGCTGAGTAAACCTGGCGAATTAGTATGATCGCGAC 3540
Db 3481 GATCCTGATGTTGATCGTGCCTAGCTGAGTAAACCTGGCGAATTAGTATGATCGCGAC 3540
Qy 3541 GATAATACAGCCATTGAGATTAATCTCAGGTGATGACAGGTTTCAATCGCTTTTCCGGGC 3600
Db 3541 GATAATACAGCCATTGAGATTAATCTCAGGTGATGACAGGTTTCAATCGCTTTTCCGGGC 3600
Qy 3601 CTTCAAAGGTGACACTATTGAGAGATAGTCAATAAAGGGTCTTTTGTGCGCAATTGT 3660
Db 3601 CTTCAAAGGTGACACTATTGAGAGATAGTCAATAAAGGGTCTTTTGTGCGCAATTGT 3660
Qy 3661 ACAATATCTCGCAAAATCCCTTGATC 3687
Db 3661 ACAATATCTCGCAAAATCCCTTGATC 3687

RESULT 6
AX122988
LOCUS AX122988 1983 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 2904 from Patent EP1108790.
ACCESSION AX122988
VERSION AX122988.1 GI:14040476
KEYWORDS
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
REFERENCE
1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
Patent: EP 1108790-A 2904 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
Location/Qualifiers
1..1983
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
BASE COUNT 410 a 508 c 558 g 507 t
ORIGIN
Query Match 31.4%; Score 1874.2; DB 6; Length 1983;
Best Local Similarity 96.68; Pred. No. 0;
Matches 1915; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
Qy 3779 ATGGACCATAGGACCTCGCGCAACGCATCCTCGGACATTTGGCGGCGAAGACACATT 3838
Db 1 ATGGACCATAGGACCTCGCGCAACGCATCCTCGGACATTTGGCGGCGAAGACACATT 60
Qy 3839 GTCCGCCCGCACACTGTGCAACCGCTTACGCGCTCGTCTCAAGACACACCAGGATG 3898
Db 61 GTCCGCCCGCACACTGTGCAACCGCTTACGCGCTCGTCTCAAGACACACCAGGATG 120
Qy 3899 GATCGCCAAAGTGTGGATGATGATCCAGATCTGAAAGGCACCTTTGAAACTGGCGGCATG 3958
Db 121 GATCGCCAAAGTGTGGATGATGATCCAGATCTGAAAGGCACCTTTGAAAGCGGTGATG 180
Qy 3959 TTCCAGATCATCTGCGGGCGAGCGATGTGATCATGTTTCAAGAACTCGATGACGCA 4018
Db 181 TTCCAGATCATCTGCGGGCGAGCGATGTGATCATGTTTCAAGAACTCGATGACGCA 240
Qy 4019 ACCTCCAAAGACATCGTGTGTCCACAGACAGCTCAAGATGTTGTGGCTAACACGCC 4078
Db 241 ACCTCCAAAGACATCGTGTGTCCACAGACAGCTCAAGATGTTGTGGCTAACACGCC 300
Qy 4079 AACTGGTTACGCGGTGTGGAAGTATTGGCGGACATTTTCGTCGCCGTGATTCACAATC 4138
Db 301 AACTGGTTACGCGGTGTGGAAGTATTGGCGGACATTTTCGTCGCCGTGATTCACAATC 360
Qy 4139 TTGGTTGGTGGCGGTGTGTCATGCTATCAACAATGTGTTGGTTCGCGAGGATGTTTC 4198
Db 361 TTGGTTGGTGGCGGTGTGTCATGCTATCAACAATGTGTTGGTTCGCGAGGATGTTTC 420
Qy 4199 GTTCCGCAATCACTGTGGAGATGTTCCCTCAGATCAGCGGTGCTGAGATGATCAAC 4258
Db 421 GTTCCGCAATCACTGTGGAGATGTTCCCTCAGATCAGCGGTGCTGAGATGATCAAC 480
Qy 4259 CTGATGCGATCTCGCGGTTCGGTCTTTCGCCAGTGTGTTGGTTTTCACCGCAACCAAG 4318
Db 481 CTGATGCGATCTCGCGGTTCGGTCTTTCGCCAGTGTGTTGGTTTTCACCGCAACCAAG 540
Qy 4319 CGTTTCGGTGGCAATGATGTTCCCTGGCGCGGATTTGGTATGCGGATGTTTCCCAAC 4378
Db 541 CGTTTCGGTGGCAATGATGTTCCCTGGCGCGGATTTGGTATGCGGATGTTTCCCGAGC 600
Qy 4379 CTGGTTAACGGCTACGACGTGGCGGCCACCATGACCGCGGCGGAAATGCCAATGTGTC 4438
Db 601 TTGGTGAACGGCTACGACGTGGCGGCCACCATGACCGCGGCGGAAATGCCAATGTGTC 660

Qy 4439 CTGTTTGGTTGGATGTTCTCAAGCTGGTTACACAGGCACCGGCTTCTCTGTCTGCTG 4498
Db 661 CTGTTTGGTTTAGATGTTGCCCAAGCCGTTTACACAGGCACCGTCTTCTCTGTCTGCTG 720
Qy 4499 GTCTCTTGGATCTGGCAACGATCGAGAGTTCCTGCAACAGCGACTCATGGGCACTGCA 4558
Db 721 GTTCTTGGATCTTGGCAACGATCGAGAGTTCCTGCAACAGCGACTCATGGGCACTGCA 780
Qy 4559 GACTTTCCTGATCACCCCACTGTTGACTCTGCTGCTTCCAGGCTTCTCTTACGTTTCT 4618
Db 781 GACTTTCCTGATCACCTCCAGTCTGAGCTGCTGCTTCCAGGATCTCTTACATTCATGCT 840
Qy 4619 ATTGCTCAGCAATCGGCTGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4678
Db 841 ATTGCTCAGCAATCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 900
Qy 4679 GATTTCGGTGGTTCCTGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4738
Db 901 GATTTCGGTGGTTCCTGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy 4739 ACTGCTCTGACCACTCTTCCGCAATTTGAGCTGGAGCTGTTCAACACAGGCTGGATCC 4798
Db 961 ACTGCTCTGACCACTCTTCCGCAATTTGAGCTGGAGCTGTTTAACACAGGCTGGATCC 1020
Qy 4799 TTCATCTTCGCAACCCCATCCATGCGCAATATCGCGCAGGCTGCGAGCTGTTTGGCAGTG 4858
Db 1021 TTCATCTTCGCAACCCCATCTATGGCTATATCGCGCAGGCTGCGGCTGTTTGGCAGTG 1080
Qy 4859 TTCCTCTAGCAAGAGTGAAGTCAAGGCTTGCAGGTGCTTCAAGTGTCTCCGCT 4918
Db 1081 TTCCTCTAGCAAGAGTGAAGTCAAGGCTTGCAGGTGCTTCAAGTGTCTCCGCT 1140
Qy 4919 GTTCTGTGTTTACAGAGCTGCGATCTTGGTGTGGAACCTTGGCTGCGCTGCGCTGCTC 4978
Db 1141 GTTCTGTGTTTACAGAGCTGCGATCTTGGTGTGGAACCTTGGCTGCGCTGCGCTGCTC 1200
Qy 4979 FACATTTGATCGGTACCGAGCTATCGGTGGCGCTTGTGATGCTGCTGCTGCTGCTGCT 5038
Db 1201 FACATTTGATCGGTACCGAGCTATCGGTGGCGCTTGTGATGCTGCTGCTGCTGCTGCT 1260
Qy 5039 GCAGTTGGTGTGGGCGCTGCGAGTCTTGGTGTGGAACCTTGGCTGCGCTGCGCTGCTC 5098
Db 1261 GCAGTTGGTGTGGGCGCTGCGAGTCTTGGTGTGGAACCTTGGCTGCGCTGCGCTGCTC 1320
Qy 5099 GTCATGTTCTGTTGGTGGGCTGAGTACCTTTGTCATCGCATTCGGCGGCGAGGATGCT 5158
Db 1321 GTCATGTTCTGTTGGTGGGCTGAGTACCTTTGTCATCGCATTCGGCGGCGAGGATGCT 1380
Qy 5159 TATGCGCTTACTTGGTTCGCGCAACCGCAGCTTATCCAGATGCAACCGCTGCTCCA 5218
Db 1381 TATGCGCTTACTTGGTTCGCGCAACCGCAGCTTATCCAGATGCAACCGCTGCTCCA 1440
Qy 5219 GTGCTTCAGGCAACCAAGCGGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5278
Db 1441 GTGCTTCAGGCAACCAAGCGGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
Qy 5279 ACCATCATCCAGGCAACCTTTGACCGGTGAAGCTATCGCACTGAGCAGGCTCAGCGATGCC 5338
Db 1501 ACCATCATCCAGGCAACCTTTGACCGGTGAAGCTATCGCACTGAGCAGGCTCAGCGATGCC 1560
Qy 5339 ATGTTTGGCAGCGGAAAGCTTGGCTGCGGCGTGGCATCGTCCCAACCAAGGCGAGTTA 5398
Db 1561 ATGTTTGGCAGCGGAAAGCTTGGCTGCGGCGTGGCATCGTCCCAACCAAGGCGAGTTA 1620
Qy 5399 GTTTCACAGCTGAGCGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5458
Db 1621 GTTTCACAGCTGAGCGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Qy 5459 GCGCAATAGGCTGAGGATGTTTCCAAATGTGGATATCTTGTGATGCACTGTTTCCGACACC 5518
Db 1681 GCGCAATAGGCTGAGGATGTTTCCAAATGTGGATATCTTGTGATGCACTGTTTCCGACACA 1740

```

Qy 5519 GTAAACCTCAACGGCAGCAGCTTTAAACCGCTGAAGAGCAGGCGGATGAAGTCAAAGCA 5578
    |||||||
Db 1741 GTAAACCTCAACGGCAGCAGCTTTAAACCGCTGAAGAGCAGGCGGATGAAGTCAAAGCA 1800
    |||||||
Qy 5579 GGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGCAGGTTATGAGGTAACC 5638
    |||||||
Db 1801 GGGAGCTGCTGTGTAATTCGATATTGATGCCATTAAAGCTGCAGGTTATGAGGTAACC 1860
    |||||||
Qy 5639 ACGCCGATGTTGTTTCGAATTAACAAGAAACCGGACCTGTAAACACTTTACGGTTTGGGC 5698
    |||||||
Db 1861 ACGCCGATGTTGTTTCGAATTAACAAGAAACCGGACCTGTAAACACTTTACGGTTTGGGC 1920
    |||||||
Qy 5699 GAAATGAAGCGGAGCAGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACA 5758
    |||||||
Db 1921 GAAATGAAGCGGAGCAGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACA 1980
    |||||||
Qy 5759 CCA 5761
    |||
Db 1981 CCA 1983

RESULT 7
BD165105
LOCUS          BD165105
DEFINITION    Novel polynucleotide.
ACCESSION     BD165105
VERSION       BD165105.1 GI:27870917
KEYWORDS      JP 2002191370-A/2904.
SOURCE        unidentified
ORGANISM      unclassified.
REFERENCE     1 (bases 1 to 1983)
AUTHORS       Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
              Yokoi,H., Tateishi,N., Senoo,A., Ikeda,M. and Ozaki,A.
TITLE         Novel polynucleotide
JOURNAL       Patent: JP 2002191370-A 2904 09-JUL-2002;
              KYOWA HAKKO KOGYO CO LTD
COMMENT       OS Corynebacterium glutamicum
              PN JP 2002191370-A/2904
              PD 09-JUL-2002
              PF 15-DEC-2000 JP 2000405096
              PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI,
              PI KEIKO OCHIAI,
              PI HARUHIKO YOKOI, NAKKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO
              PI OZAKI
              PC C12N15/09, C12N15/09, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC
              C12N1/15,
              PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P1/40, C12P13/
              PC 04, C12P13/08,
              PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC
              G01N33/566,
              PC G01N33/569, G01N33/68, G01N37/00, C12P21/08, (C12N1/21, C12P1:15),
              PC (C12N1/21, C12P1:13), (C12N1/21, C12P1:01), (C12P13/08, C12P1:15),
              PC C12N15/00,
              PC C12N5/00, C12N15/00
              CC Novel polynucleotide
              FH Key
              FT Location/Qualifiers
              FT Source
              1..1983
              /organism="Corynebacterium glutamicum".
              Location/Qualifiers
              1..1983
              /organism="unidentified"
              /mol_type="genomic DNA"
              /db_xref="taxon:32644"

BASE COUNT    410 a 508 c 558 g 507 t
ORIGIN
Query Match      31.4%; Score 1874.2; DB 6; Length 1983;
Best Local Similarity 96.6%; Pred. No. 0;
Matches 1915; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
Qy 3779 ATGGACCATATAGGACCTCGCGCAACGCGATCTCGCGACATTTGGCGGGAAGACAACATT 3838
    |||||||

```

```

Db 1 ATGGACCATATAGGACCTCGCGCAACGCGATCTCGCGACATTTGCGGGGAAGACAACATT 60
    |||||||
Qy 3839 GTGCGCGCGCACACTGTGCAACGCGTTTACGCCCTCGTGTCTCAAGACACCAAGAGTGTG 3898
    |||||||
Db 61 GTGCGCGCGCACACTGTGCAACGCGTTTACGCCCTCGTGTCTCAAGACACCAAGAGTGTG 120
    |||||||
Qy 3899 GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAAGCACCTTTTGAACCTGGCGGATG 3958
    |||||||
Db 121 GATCGCCAAAGTCTGGATGATGATCCAGATCTGAAAGCACCTTTTGAACCGGTGGTATG 180
    |||||||
Qy 3959 TTCCAGATCATGTCGCGCCAGGCGATGTGGATCATGTTTTTCAAGAACTCGATGACGCA 4018
    |||||||
Db 181 TTCCAGATCATGTCGCGCCAGGCGATGTGGATCATGTTTTTCAAGAACTCGATGACGCA 240
    |||||||
Qy 4019 ACCTCCAAAGACATCGCTGTGTCACAGACAGAGCTCAAAAGATGTTGTGGCTTAAACAGCC 4078
    |||||||
Db 241 ACCTCCAAAGACATCGCTGTGTCACAGACAGAGCTCAAAAGATGTTGTGGCTTAAACAGCC 300
    |||||||
Qy 4079 AACTGTTTCAGCCGTCGTGTAAGGATTTGCGGGACATTTTCGTCGCCCTGATCCCAATC 4138
    |||||||
Db 301 AACTGTTTCAGCCGTCGTGTAAGGATTTTCGCGGACATTTTCGTCGCCCTGATCCCAATC 360
    |||||||
Qy 4139 TTGGTTGTGGGGTCTGCTCATGGCTATCAACAATGTGTTGGTTCGCGCAGATCTGTC 4198
    |||||||
Db 361 TTGGTTGTGGGGTCTGCTCATGGCTATCAACAATGTGTTGGTTCGCGCAGATCTGTC 420
    |||||||
Qy 4199 GGTCCGCAATCACTGCTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 4258
    |||||||
Db 421 GGTCCGCAATCACTGCTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 480
    |||||||
Qy 4259 CTGATGGCATCTGCGCGCTTCGCGTCTTCGCGAGTGTGTTGGTTCACCGCAACCAAG 4318
    |||||||
Db 481 CTCATGGCATCTGCGCGCTTCGCGTCTTCGCGAGTGTGTTGGTTCACCGCAACCAAG 540
    |||||||
Qy 4319 CGTTTCGGTGGCAATGAGTTCTCGCGCGCGCATTTGTTATGGCATGTTGTTCCCAACC 4378
    |||||||
Db 541 CGTTTCGGCGCAATGAGTTCTCGCGCGCGTATTGTTGATGCGCATGTTGTTCCCGAGC 600
    |||||||
Qy 4379 CTGGTTAAGCGGTACGACGTTGCGCGCCACCATGACCGCGGGCGAATGCCATGTTGGTCC 4438
    |||||||
Db 601 TTGGTGAACGGGTACGACGTTGCGCGCCACCATGTTGCGGGCGAATGCCAATGTTGGTCC 660
    |||||||
Qy 4439 CTGTTTGGTTTGGATGTTGCTCAAGCTGTTTACCAGGCGACCGTCTCTCTGCTGTTGGTG 4498
    |||||||
Db 661 CTGTTTGGTTTAGATGTTTCCCAAGCGGTTACCAGGCGACCGTCTCTCTGCTGTTGGTG 720
    |||||||
Qy 4499 GTCCTTTGATGTTGCGCAACGATTCGAGAGTTCCTCGCAACGAGCTATGCGGCACTGCA 4558
    |||||||
Db 721 GTTTCCTTGGATTCTGCGCAACGATTCGAGAGTTCCTCGCAACGAGCTATGCGGCACTGCA 780
    |||||||
Qy 4559 GACTTCTGATCAACCCAGTGTGTTGACTCTGCTGCTCACCAGGCTTCCTTACGTTCAATTGCT 4618
    |||||||
Db 781 GACTTCTGATCACTCCAGTGTGCTGCTGCTCACCAGGATTCCTTACATTCATCGCC 840
    |||||||
Qy 4619 ATTGTTCCAGCAATGCGCTGGGTGGGTGACTTGTCTGGCACACGCTGTCAGGAGCTCTAT 4678
    |||||||
Db 841 ATTGTTCCAGCAATGCGCTGGGTGGGTGACTTGTCTGGCACACGCTGTCAGGAGCTCTAT 900
    |||||||
Qy 4679 GATTTCGGTGGTTCGACGTCGCGCTGCTTTTCGGTCTGCTTACTCACCATACTGTTATC 4738
    |||||||
Db 901 GATTTCGGTGGTTCGACGTCGCGCTGCTCTCTCGGTCTGCTTACTCACCATACTGTTATC 960
    |||||||
Qy 4739 ACTGGTCTGCACCCAGCTCTTCCCGCAATTGAGCTGAGCTGTTCAACACCGGTGGATCC 4798
    |||||||
Db 961 ACTGGTCTGCACCCAGCTCTTCCCGCAATTGAGCTGAGCTGTTTAAACAGGTTGATCC 1020
    |||||||
Qy 4799 TTCACTTTCGCAACCGCATCCATGCGCAATATCGCGAGGGTGCACATGTTTGGCAGTG 4858
    |||||||
Db 1021 TTCACTTTCGCAACCGCATCTATGGTAATATCGCGAGGGTGCAGCATGTTTGGCAGTG 1080
    |||||||
Qy 4859 TTCTTCTACGGAAGAGTGAAGAGCTCAAGGGCTTCGAGTGTCTTCAGTGTCTCCCGCT 4918
    |||||||
Db 1081 TTCTTCTACGGAAGAGTGAAGAGCTCAAGGGCTTCGAGTGTCTTCAGTGTCTCCCGCT 1140
    |||||||

```

QY	4919	GTCTTGGTATTACAGACGCTCGATCTCGGTGTGAACCTTCGCTGGCGGCTC	4978
Db	1141	GTCTTGGTATTACAGACGCTCGATCTCGGTGTGAACCTTCGCTGGCGGCTC	1200
QY	4979	TACATTGSPATCGGTACGACGCTATCGGTGGCGCTTTGATGCACTCTTTGATATCAAG	5038
Db	1201	TTCATCGGTATCGGTACGACGCTATCGGTGGCGCTTTGATGCACTCTTTAATATCAAG	1260
QY	5039	GCAGTTGCTTGGCGGCTGCAGGTTCTTGGGTGTGTTCTTATGATGCTCCAGATATG	5098
Db	1261	GCAGTTGCTTGGCGGCTGCAGGTTCTTGGGTGTGTTCTTATGATGCTCCAGATATG	1320
QY	5099	GTCAATGCTTGGTTGGCGGTAGTACCTTTGTATCGCATTCGCGCGACGCGATTGCT	5158
Db	1321	GTCAATGCTTGGTTGGCGGTAGTACCTTTGTATCGCATTCGCGCGACGCGATTGCT	1380
QY	5159	TATGGCTTACTTGGTTGGCGGCAACGACGACGATGATCCAGATGCAACGCTGCTCCA	5218
Db	1381	TATGGCTTACTTGGTTGGCGGCAACGACGATGATCCAGATGCAACGCTGCTCCA	1440
QY	5219	GTGCTTGCAGGACGACCAACGCAAGCAGACGACCGCGCAGAAATTTTCAACGATTCC	5278
Db	1441	GTGCTTGCAGGACGACCAACGCAAGCAGACGACCGCGCAGAAATTTTCAACGATTCC	1500
QY	5279	ACCATCATCCAGGACGCTTGGCGGTGAAGCTATCGCACTGACGAGCGTCAGCGATGCC	5338
Db	1501	ACCATCATCCAGGACGCTTGGCGGTGAAGCTATCGCACTGACGAGCGTCAGCGATGCC	1560
QY	5339	ATGTTTGCAGGCGAAAGCTTGGCTCAGGTGTTGGCATCGTCCCGACCAAGGGGAGCTG	5398
Db	1561	ATGTTTGCAGGCGAAAGCTTGGCTCAGGTGTTGGCATCGTCCCGACCAAGGGGAGCTG	1620
QY	5399	GTTTACCAGTGAAGCAAGATCGGTGGCTTCCGATCTGGTACGCTTCCGAGTC	5458
Db	1621	GTTTACCAGTGAAGCAAGATCGGTGGCTTCCGATCTGGTACGCTTCCGAGTC	1680
QY	5459	CGCACTAAGGCTGAGGATGTTCCCAATGTGGATATCTTGATGCACATTTGGTTTCGACAC	5518
Db	1681	CGCACTAAGGCTGAGGATGTTCCCAATGTGGATATCTTGATGCACATTTGGTTTCGACAC	1740
QY	5519	GTAACCTCAACGGCGACGACATTAACCGCTGGAAGCAAGCAGGCGGATGAAGTCAAGCA	5578
Db	1741	GTAACCTCAACGGCGACGACATTAACCGCTGGAAGCAAGCAGGCGGATGAAGTCAAGCA	1800
QY	5579	GGGAGGTGCTGTGAATTCGATATTGATGCCATTAAAGCTGCGAGTTATGAGGTAACC	5638
Db	1801	GGGAGGTGCTGTGAATTCGATATTGATGCCATTAAAGCTGCGAGTTATGAGGTAACC	1860
QY	5639	ACGCGGATGCTTTCGATTAACAAAGCAAGCGGACCTGTAACACATTAACGTTGGCG	5698
Db	1861	ACGCGGATGCTTTCGATTAACAAAGCAAGCGGACCTGTAACACATTAACGTTGGCG	1920
QY	5699	GAATTAAGCGGAGCGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACA	5758
Db	1921	GAATTAAGCGGAGCGCAACCTGCTCAACGTCGCAAGAAAGAAAGCGGTGCCAGCAACA	1980
QY	5759	CCA 5761	
Db	1981	CCA 1983	

RESULT 8
LOCUS BD094213 1656 bp DNA linear PAT 27-AUG-2002
DEFINITION Genes for heat resistant enzymes of amino acid biosynthetic pathway
derived from thermophilic coryneform bacteria.
BD094213
ACCESSION BD094213.1 GI:22639801
VERSION WO 0125447-A/71.
KEYWORDS Corynebacterium thermoaminogenes
SOURCE Corynebacterium thermoaminogenes
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE	Corynebacterineae; Corynebacteriaceae; Corynebacterium.		
AUTHORS	1 (bases 1 to 1656) Hirano, S., Nonaka, G., Matsuzaki, Y., Akiyoshi, N., Nakamura, K., Kimura, E., Osumi, T., Matsui, K., Kawahara, Y., Kurahashi, O., Nakamatsu, T. and Sugimoto, S. Genes for heat resistant enzymes of amino acid biosynthetic pathway derived from thermophilic coryneform bacteria Patent: WO 0125447-A 71 12-APR-2001; AJINOMOTO CO INC, SEIKO HIRANO, GEN NONAKA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KANAE NAKAMURA, EIICHIRO KIMURA, TSUYOSHI OSUMI, KAZUHIKO MATSUI, YOSHIO KAWAHARA, OSAMU KURAHASHI, TSUYOSHI NAKAMATSU, SHINICHI SUGIMOTO		
TITLE	Corynebacterium thermoaminogenes		
JOURNAL	PN WO 0125447-A/71 PD 12-APR-2001 PF 04-OCT-2000 WO 2000JP006913 PP 04-OCT-1999 JP 99P 282716, 01-NOV-1999 JP 99P 311147 PR 21-APR-2000 JP 00P 120687 PI SEIKO HIRANO, GEN NONAKA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KANAE PI NAKAMURA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KANAE PI EIICHIRO KIMURA, TSUYOSHI OSUMI, KAZUHIKO MATSUI, YOSHIO KAWAHARA, PI OSAMU KURAHASHI, TSUYOSHI NAKAMATSU, SHINICHI SUGIMOTO PC C12N15/60, C12N15/54, C12N15/53, C12N15/31, C12N15/56, C12N9/88, PC C12N9/12, PC C12N9/04, C07K14/34, C12N9/26, C12N13/04		
COMMENT	OS Corynebacterium thermoaminogenes PN WO 0125447-A/71 PD 12-APR-2001 PF 04-OCT-2000 WO 2000JP006913 PP 04-OCT-1999 JP 99P 282716, 01-NOV-1999 JP 99P 311147 PR 21-APR-2000 JP 00P 120687 PI SEIKO HIRANO, GEN NONAKA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KANAE PI NAKAMURA, YUMI MATSUZAKI, NAOKI AKIYOSHI, KANAE PI EIICHIRO KIMURA, TSUYOSHI OSUMI, KAZUHIKO MATSUI, YOSHIO KAWAHARA, PI OSAMU KURAHASHI, TSUYOSHI NAKAMATSU, SHINICHI SUGIMOTO PC C12N15/60, C12N15/54, C12N15/53, C12N15/31, C12N15/56, C12N9/88, PC C12N9/12, PC C12N9/04, C07K14/34, C12N9/26, C12N13/04		
FEATURES	Location/Qualifiers 1..1656 /organism="Corynebacterium thermoaminogenes" /mol_type="genomic DNA" /db_xref="taxon:161484"		
source	392 a 456 c 432 g 376 t		
BASE COUNT	392 a 456 c 432 g 376 t		
ORIGIN			

Query Match	24.4%	Score 1457;	DB 6;	Length 1656;
Best Local Similarity	93.2%;	Pred. No. 0;		
Matches 1545;	Conservative 0;	Mismatches 110;	Indels 2;	Gaps 2;
QY	2031	ACGCTCGATTCTTCAACACCATCGAAGAGGTCCCAACCGCCGCTCACCCAGGCTTGG	2090	
Db	1	ACGCCGATTCTTCAACACTATCGAAGAGTCCCAACCCAGCGCTGACCCAGGCTTGG	60	
QY	2091	GCATTTTCCCGCGCGCAAAACATCGTGTGGTGGCACTGTGTGAAGAAACCGAC-	2149	
Db	61	GTACTTTTCCCGCGCGCAAAATATCGTGTGGTGGCACTGTGTGAAGAAACCGAC	120	
QY	2150	GCCATCCGCGAAGTGTGGAAGGCCAGTGTCTTCTTCCAGAGTTCCATCTGTAG	2209	
Db	121	GCCATCCGCGAAGTGTGGAAGTCCAGTGTCTTCTTCCAGAGTTCCATCTGTCAA	180	
QY	2210	ATGCACACATGCCACCATCATCGTTGGATGAAGCAGCATATCCAGCTGGAAGCGCT	2259	
Db	181	ATGCACACACAGCCACCATCATCGTTGGATGAAGCAGCATATCCAGCTGGAAGCGCT	240	
QY	2270	GATCACTACCTCTCATCGAGCAATTAAGCTGGCTAGAACAAAGGAAGTACTGT	2329	
Db	241	GACCATACCGTCTCATGGAGCAATTAAGCTGGCTAGAACAAAGGAAGTACTGT	300	
QY	2330	GTGGGGCTATGCACACAGAACTTCCAGTTTCCGCCCTGCGTACCATGTGCTCTCGC	2389	
Db	301	GTGGGGCTATGCACACAGAACTTCCAGTTTCCGCCCTGCGTACCATGTGCTCTCGC	360	
QY	2390	AGGGCAGGCTCAATCATCCCAAGCAATGATAGTGGATGAGATACCTCCACGCTACT	2449	
Db	361	AGGGCAGGCTCAATCATCCCAAGCAATGATAGTGGATGAGATACCTCCACGCTACT	420	
QY	2450	ACCAGCAGATCCAGGTTTCCCTTTCGCACCAAAAGCGCACCGGCTGGGCTCACACCA	2509	
Db	421	ACCAGCAGATCCAGGTTTCCCTTTCGCACCAAAAGCGCACCGGCTGGGCTCACACCA	480	

QY	4558	AGACTTCCTGATCACCCAGTGTGACTCTGCTGCTCAACGGGTCTCCTTACGTTTCATTGC	4617
Db	301	AGACTTCCTGATCACCTCCAGTGTCTGCTCAACGGATTCCTTACATTCATCGC	360
QY	4618	TATTGTCCAGCAATGCGTGGTGGTGCACACGGTCTCGAGGGAGCTTA	4677
Db	361	CATTGGCCAGCAATGCGTGGTGGCGATGCTGGCACACGGTCTCAGGGAGCTTA	420
QY	4678	TGATTTCCGTGTGCTCAGTCCGGGTCTGCTTTTCGGTCTGGTCTACTCACCAATCGTTAT	4737
Db	421	TGATTTCCGTGTGCTCAGTCCGGGTCTGCTTTTCGGTCTGGTCTACTCACCAATCGTCA	480
QY	4738	CAGTGTCTGCACCACTCTCCCGCAATTGAGCTGGAGCTGTTCAACACAGGTCGATC	4797
Db	481	CAGTGTCTGCACCACTCTCTCCCGCAATTGAGCTGGAGCTGTTTAAACAGGGTGGATC	540
QY	4798	CTTCATCTTTGGCAACCGCATCCATGGCCAAATATCGCGCAGGGTGCAGCATGTTTGGCAGT	4857
Db	541	CTTCATCTTTGGCAACCGCATCTATGGCTAATATCGGCCACGGGTGCGCATGTTTGGCAGT	600
QY	4858	GTCTCTCCATAGCAAGATGAAAGCTCAAGGGCTTGCAAGTGTTCAGGTGCTCTCCCGC	4917
Db	601	GTCTCTCCATAGCAAGATGAAAGCTCAAGGGCTTGCAAGTGTTCAGGTGCTCTCCCGC	660
QY	4918	TGTTCTTGSPATTACAGAGCCTCGATCTTCGGTGTGAACCTTCGGCTCGCGTGGCGGTT	4977
Db	661	TGTTCTTGSPATTACAGAGCCTCGATCTTCGGTGTGAACCTTCGGCTCGCGTGGCGGTT	720
QY	4978	CTACATTTGATCGGTACCGCAGCTATCGGTGGCGCTTTCATTCGACTCTTTTCATATCAA	5037
Db	721	CTTCATCGGTATCGGTACCGCAGCTATCGGTGGCGCTTTCATTCGACTCTTTTCATATCAA	780
QY	5038	GGCAGTTGCTTTGGGGCTCGAGTTTCTTTGGGTGTGTTTCTATGATGCTCCAGATAT	5097
Db	781	GGCAGTTGCTTTGGGGCTCGAGTTTCTTTGGGTGTGTTTCTATGATGCTCCAGATAT	840
QY	5098	GGTCATGTTCTTGGTTTGGCGGTAGTTACCTTTGTCATCGCATTCGGCGCAGCGATTGC	5157
Db	841	GGTCATGTTCTTGGTGTGCGAGTTGTTACCTTCTTCATCGCATTCGGCGCAGCGATTGC	900
QY	5158	TTATGGCCTTTACTTGTTCGGCGAACGGCAGCATTTGTCAGATGCAACCGTGTGCTC	5217
Db	901	TTATGGCCTTTACTTGTTCGGCGAACGGCAGCATTTGTCAGATGCAACCGTGTGCTC	960
QY	5218	AGTGCCTGCAAGAACGACCAAGACCGAAGCAGACACCGCAGAAATTTTCAACAGATT	5277
Db	961	AGTGCCTGCAAGAACGACCAAGACCGAAGCAGACACCGCAGAAATTTTCAACAGATT	1020
QY	5278	CACCATCATCCAGCACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTCAGCGATGC	5337
Db	1021	CACCATCATCCAGCACCTTTGACCGGTGAAGCTATTCGCACTGAGCAGCGTCAGCGATGC	1080
QY	5338	CATGTTTGCACGCGAAAGCTTGCTCAGGTGTTTCGATCGTCCCGCACCAGGGGCGAGCT	5397
Db	1081	CATGTTTGCACGCGAAAGCTTGCTCAGGTGTTTCGATCGTCCCGCACCAGGGGCGAGTT	1140
QY	5398	GGTTTCCAGTGAACGGAAGATCGTGGTGGCTTCCCATCTGGTCAACGCTTTCGAGT	5457
Db	1141	AGTTTCTCCGGTGAAGTGGAAAGATGTGGTGGCATTCCTCATCTGGCCATGCTTTTCGAGT	1200
QY	5458	CCGCACTAAGGCTCAGGATGTTTCCCATGTGGATATCTTGATGCACATTTGTTTCGACAC	5517
Db	1201	TCGCAACCAAGCTCAGGATGTTTCCCATGTGGATATCTTGATGCACATTTGTTTCGACAC	1260
QY	5518	CGTAAACCTCAACGGCAGCACTTTAACCGCTGAAGAACGAGGGCGATGAAGTCAAGC	5577
Db	1361	AGTAAACCTCAACGGCAGCACTTTAACCGCTGAAGAACGAGGGCGATGAAGTCAAGC	1320
QY	5578	AGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAAGGCTCGAGGTTATGAGGTAA	5637
Db	1321	AGGGAGCTGCTGTGTGAATTCGATATTGATGCCATTAAAGGCTCGAGGTTATGAGGTAA	1380

Qy	5638	CAGCGCGATTTGTTTTCGAATTTACAGAAAAACCGGACCTGTATAACACTTTACGGTTTGGG	5699		
Db	1381	CAGCGCGATTTGTTTTCGAATTTACAGAAAAACCGGACCTGTATAACACTTTACGGTTTGGG	1440		
Qy	5698	CGAAATTTGAAGCGGAGCCAACTGCTCAACGTCGCAAAAGAAAGACGGTGCACGCAAC	5757		
Db	1441	CGAAATTTGAAGCGGAGCCAACTGCTCAACGTCGCAAAAGAAAGACGGTGCACGCAAC	1500		
Qy	5758	ACCATAAGTTGAACACCTTGAGTGTTCG	5784		
Db	1501	ACCATAAGTTGAACACCTTGAGTGTTCG	1527		
RESULT 11					
AX065221	AX065221	1342 bp	DNA	linear	PAT 24-JAN-2001
DEFINITION	Sequence 347 from Patent WO0100844.				
ACCESSION	AX065221				
VERSION	AX065221.1	GI:12542933			
KEYWORDS					
SOURCE	Corynebacterium glutamicum				
ORGANISM	Corynebacterium glutamicum				
REFERENCE					
AUTHORS	Pompeius M., Kroeger, B., Schroeder, H., Zeidler, O. and Haberhauer, G.				
TITLE	corynebacterium glutamicum genes encoding proteins involved in carbon metabolism and energy production				
JOURNAL	Patent: WO 0100844-A 347 (4-JAN-2001);				
FEATURES	BASE AKTTEGENSELLSCHAFT (DE)				
	Location/Qualifiers				
	1..1342				
	/organism="Corynebacterium glutamicum"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1718"				
	101..>1342				
	/note="RXA00041"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAC25850.1"				
	/db_xref="GI:12542934"				
	/translation="MHTELSRLPAYHVTPQGLNDPNMNYVDGDTLHVYXOHPDGF				
	FPAPKRTGWAHTTPLTLPQRLQWTHLPDALYDASXDLDCGSGGAVFGDGLKLFY				
	TGNLKDIDKRRATONLVEDEPTGLMGDIHRRSPKNPLIDGPASGTFPHYRDMSPID				
	GDGNKVIYGAQRENLAAGVLYRSTDLNENFEIGEITFSDAQGPAGAPDLVPGGYM				
	ECNPLFLRDEETGDLVIFCQGLDRIHDEYTHVSSDQCGYVYKLEGTFTVRL				
	RGSLELDFGEHFYAPAVAGSADWLWGMGLPAQDDHPTVAREGWHVCLTVPRKLHL				
	RNHALYQLELLPEGESGIVRSVLGSEPVPRVDIRGNISLEWDLVRLSDRGDRRAVEV				
	KPGELVADDDNTATEITAGDQ"				
BASE COUNT	312 a	373 c	359 g	298 t	
ORIGIN					
Query Match	22.1%	Score 1321.2;	DB 6;	Length 1342;	
Best Local Similarity	99.0%;	Prod. No. 0;			
Matches 1329;	Conservative	0;	Mismatches 13;	Indels 0;	Gaps 0;
Qy	2238	ATGAGCAGCAGTATCCAAAGCTGGAAACGCTGATCACTACCGTCTCATGGAGCAATTA	2297		
Db	1	ATGAGCAGCAGCAGTATCCAAAGCTGGAAACGCTGATCACTACCGTCTCATGGAGCAATTA	60		
Qy	2298	AGCTGCGCTAGAAACAAAGAAAGTACTGTGTGGGCTATGCACAGAACCTTCCAG	2357		
Db	61	AGCTGCGCTAGAAACAAAGAAAGTACTGTGTGGGCTATGCACAGAACCTTCCAG	120		
Qy	2358	TTTGGCCCTCGGTACCATGTHGACTCCCTCCGAGGCGAGGCTCAATGATCCCAACGGAAT	2417		
Db	121	TTTGGCCCTCGGTACCATGTHGACTCCCTCCGAGGCGAGGCTCAATGATCCCAACGGAAT	180		
Qy	2418	GTACGTGATGAGATACCCCTCCAGTGCTACTACAGCAGCATCCAGTTCCCGCTTCGC	2477		
Db	181	GTACGTGATGAGATACCCCTCCAGTGCTACTACAGCAGCATCCAGTTCCCGCTTCGC	240		
Qy	2478	ACCAAGCGCACCGGCTCGGGCTCAACACACACGCGCTTGACCGGACCGCAGGATTGCA	2531		

```
|||||
241 ACCAAGCGCAGCGGTGGCTCACACCACCGCGTTGACGGACCGACGATTGCA 300
2538 GTGAGCGACCTGCGCGAGCTCTTTACCGGATGATCCTATGACCTGGATGCTA 2597
|||||
301 GTGAGCGACCTGCGCGAGCTCTTTACCGGATGATCCTATGACCTGGATGCTA 360
2598 TTCGGTGGAGCGGTATTTACTGACGGACACTTAAACTTTTCTACACGGCAACTAA 2657
|||||
361 TTCGGTGGAGCGGTATTTACTGACGGACACTTAAACTTTTCTACACGGCAACTAA 420
2658 AATTGACGGAAGCGCGCCACCAACCAACCTTTCGAGTGGAGGACCCCACTGGGT 2717
|||||
421 AATTGACGGAAGCGCGCCACCAACCAACCTTTCGAGTGGAGGACCCCACTGGGT 480
2718 GATGGCGGCAATCATCGCGTTCGCTTAAATTCGCTATTCGAGCGACCCCGACGG 2777
|||||
481 GATGGCGGCAATCATCGCGTTCGCTTAAATTCGCTATTCGAGCGACCCCGACGG 540
2778 TTTACACCCCATACCGGATCCCATGATCAGCCCTGATGTTGATGTTGAAACATGT 2837
|||||
541 TTTACACCCCATACCGGATCCCATGATCAGCCCTGATGTTGATGTTGAAACATGT 600
2838 TCTTGGGGCCCAACCGGAAACCTCACCGGTGACGGTTCATACCGCTGCACAGATCT 2897
|||||
601 TCTTGGGGCCCAACCGGAAACCTCACCGGTGACGGTTCATACCGCTGCACAGATCT 660
2898 TGAACCTGGGAATTCCTCGGTGAATACCTTTGACCTCAGTATGACACACCTGGTTC 2957
|||||
661 TGAACCTGGGAATTCCTCGGTGAATACCTTTGACCTCAGTATGACACACCTGGTTC 720
2958 TGTCTCTGATCTCGTCCCGATGGCTACATGTGGGAATGCCCAACTTTTACGGTTCG 3017
|||||
721 TGTCTCTGATCTCGTCCCGATGGCTACATGTGGGAATGCCCAACTTTTACGGTTCG 780
3018 CGATGAAGAACTGGGGAAGATCTCGAGTGTGATTTCTGTCCACAGGATGGACCG 3077
|||||
781 CGATGAAGAACTGGGGAAGATCTCGAGTGTGATTTCTGTCCACAGGATGGACCG 840
3078 AATCCACGATGAGTTACTCATTACGCAAGCTCTGACGAGTGGGATATGTCTGCACAA 3137
|||||
841 AATCCACGATGAGTTACTCATTACGCAAGCTCTGACGAGTGGGATATGTCTGCACAA 900
3138 GCTTGAAGAAACGACCTTCGCGTCTTGGAGGATTCAGGAGCTGGATTTGGGCAATGA 3197
|||||
901 GCTTGAAGAAACGACCTTCGCGTCTTGGAGGATTCAGGAGCTGGATTTGGGCAATGA 960
3198 ATTCTACGACCGGATTCAGTAAACGTTCTGATGCTGCTGGCTGGGCTGGATGG 3257
|||||
961 ATTCTACGACCGGATTCAGTAAACGTTCTGATGCTGCTGGCTGGGCTGGATGG 1020
3258 GCTGCCCGCGAGGATGATCAACCAACAGTTGACAGGAGGATGGTGCACCTGCCTGAC 3317
|||||
1021 GCTGCCCGCGAGGATGATCAACCAACAGTTGACAGGAGGATGGTGCACCTGCCTGAC 1080
3318 TGTGCCCGCGAGGATGATCAACCAACAGTTGACAGGAGGATGGTGCACCTGCCTGAC 3377
|||||
1081 TGTGCCCGCGAGGATGATCAACCAACAGTTGACAGGAGGATGGTGCACCTGCCTGAC 1140
3378 GGGGAGTCCGGGGTATCAGATCTGATTAGTCTTGAACCTGTCGAGTAGACATCGG 3437
|||||
1141 GGGGAGTCCGGGGTATCAGATCTGATTAGTCTTGAACCTGTCGAGTAGACATCGG 1200
3438 AGGCAATATTCCTCGAGTGGATGGTGTCCGTTTGTCTGTGTGATGCTGTGATCG 3497
|||||
1201 AGGCAATATTCCTCGAGTGGATGGTGTCCGTTTGTCTGTGTGATGCTGTGATCG 1260
3498 TCGGAGTCCGGGGTATCAGATCTGATTAGTCTTGAACCTGTCGAGTAGACATCGG 3557
|||||
1261 TCGGAGTCCGGGGTATCAGATCTGATTAGTCTTGAACCTGTCGAGTAGACATCGG 1320
3558 GATAACTGCGAGTGTGACAG 3579
|||||
```

```
Db 1321 GATAACTGCGAGTGTGATGACAG 1342

RESULT 12
AX122989
LOCUS AX122989 1299 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 2905 from Patent EP1108790.
ACCESSION AX122989
VERSION AX122989.1 GI:14040477
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum
Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
1
Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
Patent: EP 1108790-A 2905 20-JUN-2001;
KIYWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
Location/Qualifiers
1..1299
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
BASE COUNT 287 a 363 c 350 g 299 t
ORIGIN
Query Match 21.2%; Score 1267.8; DB 6; Length 1299;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
QY 2327 TGTGTGGGGCTATGCACACAGAACTTCCAGTTTGGCCCTGGTACCATGTGACTCTC 2386
Db 2 TGTGTGGGGCTATGCACACAGAACTTTCAGTTTGGCCCTGGTACCATGTGACTCTC 61
QY 2387 CCGAGGGAGGCTCAATGATCCCAACGGAATGTAGTCGATGGAGATACCTCCACGCT 2446
Db 62 CCGAGGGAGGCTCAATGATCCCAACGGAATGTAGTCGATGGAGATACCTCCACGCT 121
QY 2447 ACTACACGAGATCCAGTTTCCCTTCGACCAAGCGCACCGGTGGCTCACACCA 2506
Db 122 ACTACACGAGATCCAGTTTCCCTTCGACCAAGCGCACCGGTGGCTCACACCA 181
QY 2507 CCACGCGCTTGACCGGACCGCAGCGATTSCAGTGGACGACCTGCCGACGCTCTTTACC 2566
Db 182 CCACGCGCTTGACCGGACCGCAGCGATTSCAGTGGACGACCTGCCGACGCTCTTTACC 241
QY 2567 CGGATGCATCTATGACCTGGATGGATCGTATTCCGTTGGAGCGGTATTACTACGCGCA 2626
Db 242 CGGATGCATCTATGACCTGGATGGATCGTATTCCGTTGGAGCGGTATTACTACGCGCA 301
QY 2627 CACTTAAACTTTTCTACACCGCAACCTTAAATTTGACGGAAGCGCGCCACCCAAA 2686
Db 302 CACTTAAACTTTTCTACACCGCAACCTTAAATTTGACGGAAGCGCGCCACCCAAA 361
QY 2687 ACCTTGTGAAGTCCAGACCCCACTGGGCTGTATGGGGGATTCATCGCGGTTCGCCTA 2746
Db 362 ACCTTGTGAAGTCCAGACCCCACTGGGCTGTATGGGGGATTCATCGCGGTTCGCCTA 421
QY 2747 AAAATCCGCTTATCGACGAGCCCGCAGCGGTTTTCACACCCCATTTACCGGATCCCATGA 2806
Db 422 AAAATCCGCTTATCGACGAGCCCGCAGCGGTTTTCACACCCCATTTACCGGATCCCATGA 481
QY 2807 TCAGCCCTGTATGGTGTGGAACATGTTTCTTGGGGCCCAACGCGAAACCTCACCG 2866
Db 482 TCAGCCCTGTATGGTGTGGAACATGTTTCTTGGGGCCCAACGCGAAACCTCACCG 541
QY 2867 GTGACGCGTTCATACCGCTCGACAGATCTTGAACCTGGGAATCTCCGGTGAATCA 2926
Db 542 GTGACGCGTTCATACCGCTCGACAGATCTTGAACCTGGGAATCTCCGGTGAATCA 601
QY 2927 CCTTTGACCTAGTGTGATGCACAACTCGTTCCTCTGCTCTGCTTCCGATGCTTACA 2986
```



```
QY 3047 TGTGATTTTCTGTCACAAAGGATGGACCGAATCCAGATGAGGTACTACTACGCAA 3106
Db 722 TGTGATTTTCTGTCACAAAGGATGGACCGAATCCAGATGAGGTACTACTACGCAA 781
QY 3107 GCTGACCAAGTGGGATATGCTGCGACAAGCTTGAAGGAACGACCTTCGCGTCTTGC 3166
Db 782 GCTGACCAAGTGGGATATGCTGCGACAAGCTTGAAGGAACGACCTTCGCGTCTTGC 841
QY 3167 GAGGATTCAGCAGCTGGATTTCCGGCCATGAATTTACGACCGCAGGTTCAGTAACG 3226
Db 842 GAGGATTCAGCAGCTGGATTTCCGGCCATGAATTTACGACCGCAGGTTCAGTAACG 901
QY 3227 GTTCTGATGCTGGCTGCTGGGCTGGATGGGCTGCCCGCAGGATGATCACCACAG 3286
Db 902 GTTCTGATGCTGGCTGCTGGGCTGGATGGGCTGCCCGCAGGATGATCACCACAG 961
QY 3287 TTGCACAGAAAGTGGGTGCACTGCTGACGTGTCGCCGCAAGTTTCATTTGCCAACC 3346
Db 962 TTGCACAGAAAGTGGGTGCACTGCTGACGTGTCGCCGCAAGTTTCATTTGCCAACC 1021
QY 3347 ACGCATCTACCAAGAGCTCTTCTCCAGAGGGGAGTCGGGGTAAATCAGATCTGTAT 3406
Db 1022 ACGCATCTACCAAGAGCTCTTCTCCAGAGGGGAGTCAGGGTAAATCAGATCTGTAT 1081
QY 3407 TAGGTTCTGAACCTGTCGAGTAGACATCGGAGGCAATATTTCCCTCGAGTGGGATGG 3466
Db 1082 TAGGTTCTGAACCTGTCGAGTAGACATCGGAGGCAATATTTCCCTCGAGTGGGATGG 1141
QY 3467 TCGTTTGTCTGTGGATCGTGATGTCGTCGGGTAGCTGAGGTAAACCTGGCGAAT 3526
Db 1142 TCGTTTGTCTGTGGATCGTGATGTCGTCGGGTAGCTGAGGTAAACCTGGCGAAT 1201
QY 3527 TAGTGATCGGACGATATACAGCATTGAGATTAAGTCTGAGGTGATGACAGGTTTCA 3586
Db 1202 TAGTGATCGGACGATATACAGCATTGAGATTAAGTCTGAGGTGATGACAGGTTTCA 1261
QY 3587 TCGTTTTCCGGCCCTCAAGGTGACACTATTGAGAGA 3625
Db 1262 TCGC-TTTCGGGCTTCAAGGTGACACTATTGAGAGA 1299

RESULT 14
AX065287 1287 bp DNA linear PAT 24-JAN-2001
LOCUS AX065287
DEFINITION Sequence 413 from Patent WO0100844.
ACCESSION AX065287
VERSION AX065287.1 GI:12542999
KEYWORDS Corynebacterium glutamicum
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in
carbon metabolism and energy production
JOURNAL Patent: WO 0100844-A 413 04-JAN-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
    Location/Qualifiers
        1..1287
            /organism="Corynebacterium glutamicum"
            /mol_type="genomic DNA"
            /db_xref="taxon:1718"
            101..1267
                /note="RXN00043"
                /codon_start=1
                /transl_table=11
                /db_xref="GI:12543000"
            /translation="MAEVHYOENAGQAVKKIEGRIVTPHGVDFGLQLENGILTELS
GERAPXNAFHELPFIVESFIDLHNEGNGAFTPTGDDQARNAAQVHREGTVM
ASVSPADALAAQVNLPLCEGLGILHEGPFINACRGANQPFIFPNPTDL
AQVIAHAGKWNISITVAPETDNLTELLDLCNAHHIIASFHTDADFDFTTSAIAAKE
```

```
BASE COUNT 325 a 349 c 326 g 287 t
ORIGIN
Query Match 20.7%; Score 1237.4; DB 6; Length 1287;
Best Local Similarity 97.6%; Pred. No. 1.4e-306;
Matches 1256; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 242 AACAGCAGGCTCAAGTCGGAAGATAATTAACCTAGATCCGTAGACATAAGACATATAC 301
Db 1 AACAGCAGGCTCAAGTCGGAAGATAATTAACCTAGATCCGTAGACATAAGACATATAC 60
QY 302 GTCCTATGCTTGTCTGGAAGAACCAATAAACCCTCAGAAAGATGCGAAGTGTGCAATTA 361
Db 61 GTCCTATGCTTGTCTGGAAGAACCAATAAACCCTCAGAAAGATGCGAAGTGTGCAATTA 120
QY 362 TCAGAAAAATCAGGTCAAGCAGTTAAAAAAATGAGGAAGATTTTCCCCCTCCG 421
Db 121 TCAGAAAAATCAGGTCAAGCAGTTAAAAAAATGAGGAAGATTTTACCCCCACGG 180
QY 422 GGTGATTGATGGCTTTCTCCAACTCGAAAGCGCATCATCAGGAACTCTCTGAGAAACC 481
Db 181 GGTGATTGATGGCTTTCTCCAACTCGAAAGCGCATCATCAGGAACTCTCTGAGAAACC 240
QY 482 AGCACCTTAAAAACGACGAGTTCCACCCGAACTCCCAAGATTTTCCGGTTTTATTGA 541
Db 241 AGCACCTTAAAAACGACGAGTTCCACCCGAACTCCCAAGATTTTCCCGAGTTTTATTGA 300
QY 542 TCTTCATTAATCACGGTGGAAACGGTGGCGCTTTCTACGGGAACGACGACGACGCGAG 601
Db 301 TCTTCATTAATCACGGTGGAAACGGTGGCGCTTTCTACGGGAACGACGACGACGCGAG 360
QY 602 GAACACCGCGCAGTATCACCCGAACTGGCAGACCGTGTATGTTGCCAGCATGTTTC 661
Db 361 GAATGCCGCGCAGTATCACCCGAACTGGCAGACCGTGTATGTTGCCAGCATGTTTC 420
QY 662 GCGCGCGGCTGACGCACCTGGCAGCGCAGGTGGAAGAACTTATTCCTTGTGTGAAGAGT 721
Db 421 GCGCGCGGCTGACGCACCTGGCAGCGCAGGTGGAAGAACTTATTCCTTGTGTGAAGAGT 480
QY 722 CTTGCTGTGGCGCATTCACCTCGAGGGCGCTTTCATCAACGCATGCGGTGTGTTGCTCA 781
Db 481 CTTGCTGTGGCGCATTCACCTCGAGGGCTTTCATCAACGCATGCGGTGTGTTGCTCA 540
QY 782 AAACCCGATTTTCAATTTTCCCGCAACCCCAACAGATCTTCCCGGGTGTATCCATGCGGG 841
Db 541 AAACCCGATTTTCAATTTTCCCGCAACCCCAACAGATCTTCCCGGGTGTATCCATGCGGG 600
QY 842 AAAAGTTGATCAAAATCGATCAGTAGCGCGGAACTGACAACTTTCTGAGCTTCT 901
Db 601 AAAAGTTGATCAAAATCGATCAGTAGCGCGGAACTGACAACTTTCTGAGCTTCT 660
QY 902 CGATCTCTGCCAGCGCACCATCATCTTCTTCCTCGGGCAGCTGATGACAGATTTTGA 961
Db 661 CGATCTCTGCCAGCGCACCATCATCTTCTTCCTCGGGCAGCTGATGACAGATTTTGA 720
QY 962 TACCACCTACCAGCGCAATTCCTTGGCTTAAAGAAAAATGTGACGCTCAGCGGTACGCA 1021
Db 721 TACCACCTACCAGCGCAATTCCTTGGCTTAAAGAAAAATGTGACGCTCAGCGGTACGCA 780
QY 1022 TTTGTTCAATGGATGCTCCGCTGCATCATAGGCTCCCGGACGCTGGCGCTTTCCT 1081
Db 781 TTTGTTCAATGGATGCTCCGCTGCATCATAGGATCCCGGACGCTGGCGCTTTCCT 840
QY 1082 TGCTGCGGCACGTCCGCGGACGCGCATATGTTAGTTGATGCCGACGCGCTGTGCTTGGC 1141
Db 841 TGCTGCGGCACGTCCGCGGACGCGCATATGTTAGTTGATGCCGACGCGCTGTGCTTGGC 900
QY 1142 CGATGGAACGTGCGATCTAGCTCGTTCACACAAAGCCCTTTTTCATCAGCGACGCTGGA 1201
Db 901 CGATGGAACGTGCGATCTAGCTCGTTCACACAAAGCCCTTTTTCATCAGCGACGCTGGA 960
```

Qy	1202	AGCGCGCGAATCCAGACGGTGAGTACATTTTGGCGGCTTTTGAACGTCACCGTCACCGA	1261	Db	61	GTCTTATGCTTGTGGAAGAGCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
Db	961	AGCGCGCGAATCCAGACGGTGAGTACATTTTGGCGGCTTTTGAACGTCACCGTCACCGA	1020	Qy	362	TCAGAAAAATGAGGTCAAAGCAGTTAAAAAAATTTAGGAGAGAAATTTGTTCCCGCCCTCGG	421	
Qy	1262	TGAGTGGCGGCTCTGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1321	Db	121	TCAGAAAAATGAGGTCAAAGCAGTTAAAAAAATTTAGGAGAGAAATTTGTTACCGCCACCG	180	
Db	1021	TGGCGTGGCGGCTCTGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1080	Qy	422	GGTGATTGATGGCTTCTCCAACTGAAACGGGATCATCAGCGAACHTCTCTGGAGAAC	481	
Qy	1322	TCAGTTGGTGGCGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1381	Db	181	GGTATTGATGGCTTCTCCAACTGAAACGGGATCATCAGCGAACHTCTCTGGAGAAC	240	
Db	1081	TCAGTTGGTGGCGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1140	Qy	482	AGCACCTAAAAACGCGAGGATTCACCGCGAACTCCCGGATTTGTTCCCGGTTTATTTGA	541	
Qy	1382	AACCGTGGCGGCTCTGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1441	Db	241	AGCACCTAAAAACGCGAGGATTCACCGCGAACTCCCGGATTTGTTCCCGGTTTATTTGA	300	
Db	1141	AACCGTGGCGGCTCTGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1200	Qy	542	TCTTATATATCAGCGTGGAAACGGTGGCGGCTTCTCAGGGAAGCGAGGACCGCGAG	601	
Qy	1442	AAATTTTGTGCTTTGACTCAACGCGCGAGGTGCAAAAGSTCCATTTAGTGCATCAAGT	1501	Db	301	TCTTATATATCAGCGTGGAAACGGTGGCGGCTTCTCAGGGAAGCGAGGACCGCGAG	360	
Db	1201	AAATTTTGTGCTTTGACTCAACGCGCGAGGTGCAAAAGSTCCATTTAGTGCATCAAGT	1260	Qy	602	GAACCGCGCAGTATCAGCGGAAATGCGCAGACCGTGTATGTTGCCAAGCATGTTTC	661	
Qy	1502	ACTTTAAGTAGGAGTAACATATCCTG	1528	Db	361	GAATCCGCGCAGTATCAGCGGAAATGCGCAGACCGTGTATGTTGCCAAGCATGTTTC	420	
Db	1261	AAATTTAATACGAGCAAAACTTTCCTG	1287	Qy	662	GGCGCGGCTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTGTGTGAGAGAGT	721	
RESULT 15				Db	421	GGCGCGGCTGACGCACTGGCAGCGAGGTGGAAACCTTATTCCTGTGTGAGAGAGT	480	
LOCUS	AX065289	1287 bp	DNA	linear	Qy	722	CCTGTGTGCGGCACTTCACTCGAGGCGCTTTCATCAACGCAATGCGGTGTGTGCTCA	781
DEFINITION	Sequence 415 from Patent WO0100844.			Db	481	CCTGTGTGCGGCACTTCACTCGAGGCTTTCATCAACGCAATGCGGTGTGTGCTCA	540	
ACCESSION	AX065289			Qy	782	AAACCGGATTTCAATTTTCCCGGCAACCCACAGATCTTCCCGGGTGATCCATGCGGG	841	
VERSION	AX065289.1	GI:12543001		Db	541	AAACCGGATTTCAATTTTCCCGGCAACCCACAGATCTTCCCGGGTGATCCATGCGGG	600	
KEYWORDS	Corynebacterium glutamicum			Qy	842	AAAAGTTTGGATCAATTCGATCAGTAGCGCGGAAACTGACATCTTCTGAGGCTCT	901	
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Corynebacteriales; Corynebacterium.			Db	601	AAAAGTTTGGATCAATTCGATCAGTAGCGCGGAAACTGACATCTTCTGAGGCTCT	560	
ORGANISM	Pompejus, M., Kroeger, B., Schroeder, H., Zeider, O. and Haberhauer, G..			Qy	902	CGATCTCTGGCAGCGCACACATCATCTGCTTCTTCCGCGCACACATGATGAGATTTGA	961	
REFERENCE	corynebacterium glutamicum genes encoding proteins involved in carbon metabolism and energy production			Db	661	CGATCTCTGGCAGCGCACACATCATCTGCTTCTTCCGCGCACACATGATGAGATTTGA	720	
AUTHORS	Patent: WO 0100844-A 415 04-JAN-2001;			Qy	962	TACCACCTACCGCAATTCCTTGGCTAAAGAGAAATAATGTACGCTCACGGCTACGCA	1021	
TITLE	BASF AKTIENGESSELLSCHAFT (DE)			Db	721	TACCACCTACCGCAATTCCTTGGCTAAAGAGAAATAATGTACGCTCACGGCTACGCA	780	
JOURNAL	Location/Qualifiers			Qy	1022	TTTGTTCATTCGATCGATCGCTCCGCTGATCATAGGGCTCCCGGAGCGGTGGGCTTTGCT	1081	
FEATURES	1..1287			Db	781	TTTGTTCATTCGATCGATCGCTCCGCTGATCATAGGGATCCCGGAGCGGTGGGCTTTGCT	840	
source	/organism="Corynebacterium glutamicum"			Qy	1082	TGCTGGGCGACGTGGCGGAGCGCATATGTTGAGTTGATCGCGAGCGGTGCATTTGGC	1141	
	/mol_type="genomic DNA"			Db	841	TGCTGGGCGACGTGGCGGAGCGCATATGTTGAGTTGATCGCGAGCGGTGCATTTGGC	900	
CDS	/db_xref="taxon:1718"			Qy	1142	CGATGAAACGGTGCATCTAGCTCGTTCACAAACGCTTTTTCATCACGAGCGCATGGA	1201	
	101..1267			Db	901	CGATGAAACGGTGCATCTAGCTCGTTCACAAACGCTTTTTCATCACGAGCGCATGGA	960	
	/note="FRXA00043"			Qy	1202	AGCGCGCGGAATGCCAGACGGTGGTATGAGCTTATGCGAGCGACCTCCACACCTC	1261	
	/codon_start=1			Db	961	AGCGCGCGGAATGCCAGACGGTGGTATGAGCTTATGCGAGCGACCTCCACCGTACCGA	1020	
	/transl_table=1			Qy	1262	TGAGTGGCGGCTGCGGAGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1321	
	/protein_id="CAC25884.1"			Db	1021	TGAGTGGCGGCTGCGGAGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1080	
	/db_xref="GI:12543002"			Qy	1322	TCAGTTGGTGGCGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1381	
	/translation="MAEVVHYQENAGQVKKIEGRIVTPHGVIDGFLQLENGITELS			Db	1081	TCAGTTGGTGGCGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1140	
	GEPAKNGAPHELPITVPSFDLHNHGGGAPPTGTQDQARNAAQYHREHGTVM			Qy	1382	AACCGTGGCGGCTCTGGCGATGGCGGCGCATCGCGGCGCATCGCGGCGCACGACACATAGCGAG	1441	
	ASMSYAPADALAAQVENLIPCEGLICGTHLESPFINACRQGNQDFPFNGNPDL			Db	301	AAATTTTGTGCTTTGACTCAACGCGCGAGGTGCAAAAGSTCCATTTAGTGCATCAAGT	1260	
	AQVTHAGKWLKSTVAPEDNLTELLOCAHHIIASFQHTDADFHTTSALIAAKE			Qy	242	AAATTTTGTGCTTTGACTCAACGCGCGAGGTGCAAAAGSTCCATTTAGTGCATCAAGT	1287	
	KNVTITTHFNAMPPLHHRDPGSGVALLAARAGDAYVELIADGVHLADGTVDLARS			Db	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
	NNAFFITIDAEAGMPDGEYILGVINVTVDGVARLRDGAIRAGTSTLASQFVHYHR			Qy	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
	RGMTLIDATLHTSTVAAKILGLSDHEIVKSNPNVFNVDNSNGQLQVHLDHQVI			Db	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
BASE COUNT	325 a 349 c 326 g 287 t			Qy	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
ORIGIN				Db	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
Query Match	20.7%; Score 1237.4; DB 6; Length 1287;			Qy	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
Best Local Similarity	97.6%; Pred. No. 1.4e-306;			Db	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	
Matches 1256; Conservative	0; Mismatches 31; Indels 0; Gaps 0;			Qy	302	STCTATGCTTCTGCTGGAAGGACCAATAAOCCTCAGAAAGATGGCAGAGTGGTGCAATTA	120	

Db 1141 AACCGTCGCCCAAAATTCGGAATTAGCGATCACGAAATCGTTAAATCCAAACCCCTGT 1200
Qy 1442 AAATTTTGTGGTCTTTGACTCAAAAGGCCAGGTGCAAAAGGTCCATTAGGTCAATCAAGT 1501
Db 1201 AAATTTTGTGGTCTTTGACTCAAAAGGCCAGGTGCAAAAGGTCCATTAGGTCAATCAAGT 1260
Qy 1502 ACTTTAAGTACGAGTAAACTATCCTG 1528
Db 1261 AATTTAATACGAGCAAAACTTTCCTG 1287

Search completed: September 30, 2003, 23:16:27
Job time : 14075 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 15:17:02 ; Search time 928 seconds
(without alignments)
17363.088 Million cell updates/sec

Title: US-10-019-284A-1

Perfect score: 5969
Sequence: 1 agtcctgcagccaccatt.....tgatatattgcgcgcgtgaa 5969

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

```

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB ID	Description
1	5969	100.0	5969 22 AAF32543	Brevibacterium lac
2	5656.2	94.8	349980 22 AAH68533	C glutamicum codin
3	3687	61.8	6911 17 AAT34177	Coryneform sucrose
4	1874.2	31.4	1983 22 AAH67869	C glutamicum codin
5	1457	24.4	1656 22 AAF97497	Corynebacterium th
6	1412.6	23.7	1527 22 AAF31528	C.glutamicum phosph
7	1412.6	23.7	1527 22 AAB55346	DNA encoding C. gl
8	1321.2	22.1	1342 22 AAF71533	Corynebacterium gl

9	1267.8	21.2	1299	22	AAH67870	C glutamicum codin
10	1237.4	20.7	1287	22	AAF71566	Corynebacterium gl
11	1237.4	20.7	1287	22	AAF71567	Corynebacterium gl
12	1110.4	18.6	1152	22	AAH67872	C glutamicum codin
13	1049.8	17.6	1109	22	AAF31529	C.glutamicum phosph
14	762	12.8	882	22	AAF71571	Corynebacterium gl
15	671.8	11.3	759	22	AAH67871	C glutamicum codin
16	474.2	7.9	479	22	AAF87433	Corynebacterium th
17	466.6	7.8	498	22	AAF87432	Corynebacterium th
18	232.2	3.9	2913	22	AAH54100	S. epidermidis gen
19	215.2	3.6	1059	22	AAF71769	Corynebacterium gl
20	215.2	3.6	1059	23	AAH96088	C. glutamicum gene
21	192	3.2	975	24	ABK73538	Bacillus lichenifo
22	177.6	3.0	30246	18	AAV74367	Staphylococcus aur
23	174.2	2.9	1380	24	ABK73394	Bacillus lichenifo
24	173.8	2.9	2127	25	ABZ42306	Streptococcus pneu
25	173.8	2.9	2162598	25	ABS56454	Streptococcus pneu
26	172.2	2.9	1965	25	ABX07668	S. pneumoniae type
27	172.2	2.9	3895	19	AAV52334	Streptococcus pneu
28	163.6	2.7	1884	23	AAH55732	Streptococcus pneu
29	161	2.7	1390	22	AAH67868	C glutamicum codin
30	157	2.6	11330	23	AAH59544	Propionibacterium
31	154.2	2.6	9769	19	AAV52163	Streptococcus pneu
32	153	2.6	5840	20	AAH12968	Enterococcus faeca
33	153	2.6	5840	24	ABQ98763	Enterococcus faeca
34	139.2	2.3	3037	24	ABQ70786	Listeria monocytog
35	139	2.3	732	24	ABN91814	Staphylococcus epi
36	137.8	2.3	1917	24	ABN67846	Streptococcus poly
37	137.8	2.3	2155561	24	ABN71527	Streptococcus poly
38	137.6	2.3	2944528	24	ABA03041	Listeria monocytog
39	133.6	2.2	684707	24	ABQ67196	Listeria innocua c
40	133.6	2.2	3011208	24	ABQ69245	Listeria innocua D
41	122.2	2.0	567	20	AAH13601	Enterococcus faeca
42	122.2	2.0	567	24	ABH99396	Enterococcus faeca
43	118.4	2.0	936	22	AAH67873	C glutamicum codin
44	117.4	2.0	1185	25	ABZ42298	Streptococcus pneu
45	116	1.9	1098	24	ABK73443	Bacillus lichenifo

ALIGNMENTS

RESULT 1

AAF32543

ID AAF32543 standard; DNA; 5969 BP.

XX AAF32543;

AC

XX

DT 20-APR-2001 (first entry)

DE

XX

DE Brevibacterium lactofermentum sucrose PTS enzyme II DNA SEQ ID NO:1.

XX

XX Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;

KW Phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;

KW Coryneform bacterium; phosphoenolpyruvate-sugar transport system; ds.

XX

OS Brevibacterium lactofermentum.

XX

XX Key Location/Qualifiers

FT CDS 3779..5764

FT /*tag= a

FT /product= "sucrose PTS enzyme II"

XX

PN WO200102584-A1.

XX

PD 11-JAN-2001.

XX

XX 30-JUN-2000; 2000WO-JP04348.

XX

XX 02-JUL-1999; 99JP-0189512.

XX

PA (AJIN) AJINOMOTO CO INC.

XX

PI	Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;	
XX	WPI: 2001-138150/14.	
DR	P-PSDB; AAB69080.	
XX		
XX	Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme	
PT	II obtained by cassette ligation-mediated amplification of downstream	
PT	domain of coryneform bacterium sucrose gene, with sucrose-binding	
PT	activity -	
XX		
XX		
PS	Claim 3; Page 22-29; 45pp; Japanese.	
XX		
CC	The present sequence encodes the Brevibacterium lactofermentum sucrose	
CC	PTS (phosphoenolpyruvate:carboxylate phosphotransferase system or	
CC	phosphoenolpyruvate-sugar transport system) enzyme II, which has	
CC	sucrose-binding activity. A coryneform bacteria produced with the	
CC	sucrose PTS enzyme II gene can have more efficient sugar uptake, and	
CC	improved amino-acid and nucleic acid productivity. The sucrose PTS gene	
CC	and it's disrupted gene, such as one without the sucrose PTS function,	
CC	can be used to produce new breeds of coryneform bacterial strains to	
CC	uptake sugar more efficiently e.g. glucose only or and sucrose, and can	
CC	have improved amino-acid and nucleic acid productivity.	
XX		
SQ	Sequence 5969 BP; 1388 A; 1575 C; 1578 G; 1428 T; 0 other;	
	Query Match 100.0%; Score 5969; DB 22; Length 5969;	
	Best Local Similarity 100.0%; Pred No. 0;	
	Matches 5969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 AGTCCGTGACGCCACACATTGATGTGGTGTACCGAGCTTGGGAGGTTTCTACATCT 60	
Db	1 AGTCCGTGACGCCACACATTGATGTGGTGTACCGAGCTTGGGAGGTTTCTACATCT 60	
Qy	61 ACCTCCCGCGCGTGGAGTGGGTCATTACGGTGGGATCACGCCGCGTGAAGTTGG 120	
Db	61 ACCTCCCGCGCGTGGAGTGGGTCATTACGGTGGGATCACGCCGCGTGAAGTTGG 120	
Qy	121 GAACCCATGTGTTCCTTGTGGTTGAGGAACGAGTGCGGGTGAGAAGTTTTCAGATG 180	
Db	121 GAACCCATGTGTTCCTTGTGGTTGAGGAACGAGTGCGGGTGAGAAGTTTTCAGATG 180	
Qy	181 TCTGCAGTTTAAAGTATCATCATCAGCTTGGAGGCTGGAAGGCTGAGTATCAGTAGACCTG 240	
Db	181 TCTGCAGTTTAAAGTATCATCATCAGCTTGGAGGCTGGAAGGCTGAGTATCAGTAGACCTG 240	
Qy	241 CAACAGCAGCCCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAAGACATCAT 300	
Db	241 CAACAGCAGCCCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAAGACATCAT 300	
Qy	301 CGTCTATGCTTGTGGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGCAATT 360	
Db	301 CGTCTATGCTTGTGGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGCAATT 360	
Qy	361 ATCAAGAAATGCAAGTCAAGCAGTTAAAAAATTGAGGAAGAATTGTCCCCCTCG 420	
Db	361 ATCAAGAAATGCAAGTCAAGCAGTTAAAAAATTGAGGAAGAATTGTCCCCCTCG 420	
Qy	421 GGGTATGATGGTTTCTCCAACTCGAAACGGCATCATCAGGAACCTCTCTGGAGAAC 480	
Db	421 GGGTATGATGGTTTCTCCAACTCGAAACGGCATCATCAGGAACCTCTCTGGAGAAC 480	
Qy	481 CAGCACCTAAAGCGCAGGATCCACCCGAACTCCCGAGATTGTCGGGTTTATTG 540	
Db	481 CAGCACCTAAAGCGCAGGATCCACCCGAACTCCCGAGATTGTCGGGTTTATTG 540	
Qy	541 ATCTTTCATATCAGCGTGGAAACGGTGGCGGTTTCTACGGGAACCGCAGGACCGGA 600	
Db	541 ATCTTTCATATCAGCGTGGAAACGGTGGCGGTTTCTACGGGAACCGCAGGACCGGA 600	
Qy	601 GGAACACCGCGCAGTATCACCGGACATGGCAGCCGTGATGTGCCAAGCATGGTTT 660	
Db	601 GGAACACCGCGCAGTATCACCGGGAAGTGTATTCAGAACTGCAAGGCACTTCTGTGG 1740	
Qy	1741 ATGAATACGTGGGACTAAACCCCTGACGATGAAACAGCTACTTTTAAACCACTTCGAAA 1800	
Qy	661 CGGCGCGGCTGACGCACTGGCAGCGAGGTGGAACCTTATTCCTTTGTGTGAAGAG 720	
Db	661 CGGCGCGGCTGACGCACTGGCAGCGAGGTGGAACCTTATTCCTTTGTGTGAAGAG 720	
Qy	721 TCTGCTGTGGGCAATTCACCTCGAGGCGCTTTCATCAACGCAATGCGGTGTGCTC 780	
Db	721 TCTGCTGTGGGCAATTCACCTCGAGGCGCTTTCATCAACGCAATGCGGTGTGCTC 780	
Qy	781 AAAACCGGATTTTATTTTCCCGCAACCAACAGATCTTTCGCGGCTGATCCATCGG 840	
Db	781 AAAACCGGATTTTATTTTCCCGCAACCAACAGATCTTTCGCGGCTGATCCATCGG 840	
Qy	841 GAAAGGTTGGATCAATTCAGATGACAGTACGCGCGGAACTGAACTTCTGAGCTTC 900	
Db	841 GAAAGGTTGGATCAATTCAGATGACAGTACGCGCGGAACTGAACTTCTGAGCTTC 900	
Qy	901 TCGATCTCTGGGCGAGCGCACCATCATTTGCTTCTCGGGCACACATGATCAGATT 960	
Db	901 TCGATCTCTGGGCGAGCGCACCATCATTTGCTTCTCGGGCACACATGATCAGATT 960	
Qy	961 ATACCACTACACAGCGCAATTCCTTGGCTAAAGAGAAAATGTACGGTACAGGCTACG 1020	
Db	961 ATACCACTACACAGCGCAATTCCTTGGCTAAAGAGAAAATGTACGGTACAGGCTACG 1020	
Qy	1021 ATTTGTTCAATCGGATGCTCGCTGCATCATAGGGTCCCGGAGCGTGGCGCTTTCG 1080	
Db	1021 ATTTGTTCAATCGGATGCTCGCTGCATCATAGGGTCCCGGAGCGTGGCGCTTTCG 1080	
Qy	1081 TTGCTGCGGACGTCGCGGGACGATATGTTGAGTTGATCGCGAGCGGTGATTTGG 1140	
Db	1081 TTGCTGCGGACGTCGCGGGACGATATGTTGAGTTGATCGCGAGCGGTGATTTGG 1140	
Qy	1141 CGGATGGAACGGTGCATCTAGTCTGCTTCCAAACAGCGCTTTTTCATCACGACGCCAT 1200	
Db	1141 CGGATGGAACGGTGCATCTAGTCTGCTTCCAAACAGCGCTTTTTCATCACGACGCCAT 1200	
Qy	1201 AAGCGCGCGAATGCCAGCGTGAATACATTTTGGGCGTTTGAACGTCACCGTCAAG 1260	
Db	1201 AAGCGCGCGAATGCCAGCGTGAATACATTTTGGGCGTTTGAACGTCACCGTCAAG 1260	
Qy	1261 ATGGAGTGGCGCTGCGCGATGCGCGGCGCATCGCGGGGACACAGCACATAGCGA 1320	
Db	1261 ATGGAGTGGCGCTGCGCGATGCGCGGCGCATCGCGGGGACACAGCACATAGCGA 1320	
Qy	1321 GTCAGTTGCTGCACACGTCGCGAGGGTATGACGCTTATCGACGCGACCTTCCACACCT 1380	
Db	1321 GTCAGTTGCTGCACACGTCGCGAGGGTATGACGCTTATCGACGCGACCTTCCACACCT 1380	
Qy	1381 CAACCGTGGCGCTAAAATTCCTCGGTCCTGGGATCAGCAATCGCTAAATCCACCCCTG 1440	
Db	1381 CAACCGTGGCGCTAAAATTCCTCGGTCCTGGGATCAGCAATCGCTAAATCCACCCCTG 1440	
Qy	1441 CAAATTTTGTGCTCTTTGACTCAACGCGCGAGGTGCAAAAGTCCATTAGGTCAATCA 1500	
Db	1441 CAAATTTTGTGCTCTTTGACTCAACGCGCGAGGTGCAAAAGTCCATTAGGTCAATCA 1500	
Qy	1501 TACTTTAGTAGAGTAAACTATCCTGATTTTAAAGAGTCCCAACCATGGAATCACTA 1560	
Db	1501 TACTTTAGTAGAGTAAACTATCCTGATTTTAAAGAGTCCCAACCATGGAATCACTA 1560	
Qy	1561 TCTGCAAAAGCAGCAAGAGTGGCAAGAGTTCAGTCTCTATTCGACCCCTTCGCCA 1620	
Db	1561 TCTGCAAAAGCAGCAAGAGTGGCAAGAGTTCAGTCTCTATTCGACCCCTTCGCCA 1620	
Qy	1621 ACAAGGGTGAACCTTTGGGGTTCGACAGGATCCTCACACTGAGTACCTACCAAGAGC 1680	
Db	1621 ACAAGGGTGAACCTTTGGGGTTCGACAGGATCCTCACACTGAGTACCTACCAAGAGC 1680	
Qy	1681 TCATTTCGATGTATGAAGCTGGGAGTGTCAATCAAGAACTGAAGCACTTCTGTGG 1740	
Db	1681 TCATTTCGATGTATGAAGCTGGGAGTGTCAATCAAGAACTGAAGCACTTCTGTGG 1740	
Qy	1741 ATGAATACGTGGGACTAAACCCCTGACGATGAAACAGCTACTTTTAAACCACTTCGAAA 1800	

C glutamicum coding sequence fragment SEQ ID NO: 7058.

DE C glutamicum coding sequence fragment SEQ ID NO: 7058.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159182.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX

XX Disclosure; SEQ ID NO: 7058; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Corynebacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of corynebacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Corynebacterium, and identifying a homologue of a gene derived

CC from corynebacterium. Corynebacterium bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX

XX Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;

XX

Query Match 94.8%; Score 5656.2; DB 22; Length 349980;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

QY 1 AGTCGCGCGAGCCACCATGATGCTGGTGCACCGAGCTGCGGAGGCTTTCTACATCT 60

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 AGCTCCCGTGGCGTGGAGTGGGTCATTACGGGTGGGATCAGCCGGTGAAGTTGCG 120

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115525 ACACCTGCGCGTGGAGTGGGTCATTACGGGTGGGATCAGCCGGTGAAGTTGCG 115466

QY 121 GAACCCATGGTTCCTTGTGGTGTAGGACAGTGGCGGTGAGAGTTTCAAGTG 180

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115465 GAACCCATGGTTCCTTGTGGTGTAGGACAGTGGCGGTGAGAGTTTCAAGTG 115406

QY 181 TCTCAGTCTTTTAAGTTATGTCATCATCAGCTTGAAGGCTGAGGTAATTCAGTAGACCTG 240

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115405 TCTCAGTCTTTTAAGTTATGTCATCATCAGCTTGAAGGCTGAGGTAATTCAGTAGACCTG 115346

QY 241 CAACAGAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAGACATATA 300

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115345 CAACAGAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAGACATATA 115286

QY 301 CGTCTCTGCTGGTGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGATT 360

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115285 COTCCTATGCTTGTGGGAAGGAAATAACCTCAGAAAGATGGCAGAGTGGTGATT 115226

QY 361 ATCAAGAAATGAGTCAAGCAGTTAAATAAATTTGAGGAAAGATTTGCCCCCTCG 420

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115225 ATCAAGAAATGAGTCAAGCAGTTAAATAAATTTGAGGAAAGATTTGTTACCCCTCAG 115166

QY 421 GGGTGTATGATGGCTTCTCAACTCGAAACGGCATCATCAGCAACTCTCTGAGAAC 480

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115165 GGGTGTATGATGGCTTCTCAACTCGAAACGGCATCATCAGCAACTCTCTGAGAAC 115106

QY 481 CAGCACTAAAAACGAGGATTCCACCCCGAACTCCCGACGATTTCCCGGTTTATTG 540

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115105 CAGCACTAAAAACGAGGATTCCACCCCGAACTCCCGACGATTTCCCGGTTTATTG 115046

QY 541 ATCTTCATTAATCAGGTTGAAACGGTGGCGGCTTTCCCTACGGGAACGAGGACGAGCGA 600

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 115045 ATCTTCATTAATCAGGTTGAAACGGTGGCGGCTTTCCCTACGGGAACGAGGACGAGCGA 114986

QY 601 GGAACACCCGCGAGTATCACCGCGAACATGGCAGCAGCTGATGTTGGCAAGCATGGTTT 660

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114985 GGAATGCCCGCAGTATCACCGCGAACATGGCAGCAGCTGATGTTGGCAAGCATGGTTT 114926

QY 661 CGCGCGCGCTGACGCACTGGCAGCGAGTGGGAAACCTTATTCCTTGTGTGAGAGG 720

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114925 CGCGCGCGCTGACGCACTGGCAGCGAGTGGGAAACCTTATTCCTTGTGTGAGAGG 114866

QY 721 TCTCTGTGCGCGCATTCACCTCGAGGGGCTTTTCATCAACGATGCGGCTTGTGTGCTC 780

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114865 GCTGTGTGGCGCATTCACCTCGAGGGTCTTTTCATCAACGATGCGGCTTGTGTGCTC 114806

QY 781 AAAACCGGATTCATTTTCCCGCAACCCCAACAGATCTTCCCGGGTGTATCCATCGG 840

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114805 AAAACCGGATTCATTTTCCCGCAACCCCAACAGATCTTCCCGGGTGTATCCATCGG 114746

QY 841 GAAAAGTTGGATCAAAATCGATCAGTACGCGCGGAACTGACATCTTCTGAGCTTC 900

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114745 GAAAAGTTGGATCAAAATCGATCAGTACGCGCGGAACTGACATCTTCTGAGCTTC 114686

QY 901 TCGATCTGCGGAGCGACACATCATTTCTTCTCGGACACTGATGAGATTTG 960

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114685 TCGATCTGCGGAGCGACACATCATTTCTTCTCGGACACTGATGAGATTTG 114626

QY 961 ATACCACTACCGAGCGCAATTTGGCTTAAAGAAATAATGTGCGGTGACGGTACG 1020

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114625 ATACCACTACCGAGCGCAATTTGGCTTAAAGAAATAATGTGCGGTGACGGTACG 114566

QY 1021 ATTTGTTCAATGCGATGCTCCGCTGCATCATAGGCTCCCGGAGCGTGGCGCTTGC 1080

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114565 ATTTGTTCAATGCGATGCTCCGCTGCATCATAGGCTCCCGGAGCGTGGCGCTTGC 114506

QY 1081 TTGCTCGGCACGTGCGGGGAGCGCATATGTTGAGTTGATCGCGACGCGTGCATTTGG 1140

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114505 TTGCTCGGCACGTGCGGGGAGCGCATATGTTGAGTTGATCGCGACGCGTGCATTTGG 114446

QY 1141 CGATGGAACGGTGCATCTAGTCTGTTCCAAACAGCCCTTTTTCATCAGGAGCGCATGG 1200

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114445 CGATGGAACGGTGCATCTAGTCTGTTCCAAACAGCCCTTTTTCATCAGGAGCGCATGG 114386

QY 1201 AAGCCCGCGAATGCCAGACGGTGAATATTTTGGGCGCTTTTGAACGTCACCGTCACCG 1260

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114385 AAGCCCGCGAATGCCAGACGGTGAATATTTTGGGCGCTTTTGAACGTCACCGTCACCG 114326

QY 1261 ATGAGTCCCGCTCTGCGCGATGGCGGCGCATCGCCGGGGGACACGACATAGCGA 1320

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114325 ATGGCTCGCGCTCTGCGCGATGGCGGCGCATCGCCGGGGGATACGAGCGATAGCGA 114266

QY 1321 GTCAGTCTGTCACACAGCTGCGCAGGGGTATGACCTTATCGACGCGCCCTCCACACT 1380

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114265 GTCAGTCTGTCACACAGCTGCGCAGGGGTATGACCTTATCGACGCGCCCTCCACACT 114206

QY 1381 CAACCGCTCGCGCTAAATTTCTCGGCTTTGGGCGATCACGAAATCGCTAATCAACCGCTG 1440

DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 114205 CAACCGCTCGCGCTAAATTTCTCGGCTTTGGGCGATCACGAAATCGCTAATCAACCGCTG 114146

QY	1441	CRATTTTGTGCTTTGACTCAACGGCCAGGTGCARAAAGTTCATTTAGTCAATCAAG	1500	QY	2519	CCGACCGCAGCGATTGAGTGGAGCGACCTGCCCGAGGCTCTTTACCCGGATGCATCCT	2578
DB	114145	TAAATTTGTGCTTTGACTCAACGGCCAGTTCACAGCACTCCATTTAGACCATCAAG	114086	DB	113065	CCGACCGCAGCGATTGAGTGGAGCGACCTGCCCGAGGCTCTTTACCCGGATGCATCCT	113006
QY	1501	TACTTTAAAGTACGAGTAAACTATCCTGAT-TTTAAAGGAGTCCCAACCATGGAATCACT	1559	QY	2579	ATGACCTGGATGGATGCTATTCCCGTGGAGCGCTATTACTAGCGGCACACTTAAACTTT	2638
DB	114085	TAAATTTAAATACGAGCAAAACTTTCCTGATAATAAAAGGAGTCCGACCATGGACATCATC	114026	DB	113005	ATGACCTGGATGGATGCTATTCCCGTGGAGCGCTATTACTAGCGGCACACTTAAACTTT	112946
QY	1560	ATCTGCAAGACGAGCAAGAAGTCGCAAGAGTTCGAGTCTTAATCCGACCCCTTCGCC	1619	QY	2639	TCTACACCGGCAACCTTAAATTTGACGGAAAGCGCGCGCCACCCAAAACCTTTGTCGAAG	2698
DB	114025	ATCTGCAAGACGAGCAAGAAGTCGCAAGAGTTCGAGTCTTAATCCGACCCCTTCGCC	113966	DB	112945	TCTACACCGGCAACCTTAAATTTGACGGCAAGCGCGCGCCACCCAAAACCTTCGCGAAG	112886
QY	1620	AACAAGGGTGGAACTTGGGGCTTGCAACAGAGTCTCCACCTAGTACCTACCAAGAG	1679	QY	2699	TCGAGGACCCAACTGGGCTGATGGCGCATTCATCGCGTTCGCCCTAAATAATCCGCTTA	2758
DB	113965	ACTAAGGGCGGAACCTTGGGGCTTGCAACTGGATCTCACCCTTGAGCAGCTTACCAAGAG	113906	DB	112885	TCGAGGACCCAACTGGGCTGATGGCGCATTCATCGCGTTCGCCCTAAATAATCCGCTTA	112826
QY	1680	CTCATTCGATGATGAAGCTGGGGAAGTGTCAATCAAGACTGCAAGGCATCTTGTTG	1739	QY	2759	TCGAGGACCCCGCAGCGGTTTACACCCCATTTACCGCGATPCCCATGATCAGCCCTGATG	2818
DB	113905	CTCATTCGATGATGAAGCTGGGGAAGTGTCAATCAAGACTGCAAGGCATCTTGTTG	113846	DB	112825	TCGAGGACCCCGCAGCGGTTTACACCCCATTTACCGCGATPCCCATGATCAGCCCTGATG	112766
QY	1740	GATGAATACGTGGGACTAACCCGTGACGATGAACAGACTACTTTAAACCATTCGCRAA	1799	QY	2819	GTGATGGTTGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCGGTGCACGCTTC	2878
DB	113845	GATGAATACGTGGGATTAACGCCGACGATGAACAGACTACTTTAAACCATTCGTAAA	113786	DB	112765	GTGATGGTTGAACATGTTCTTGGGGCCCAACCGGAAACCTCACCGGTGCACGCTTC	112706
QY	1800	GAGTTCACATGACACATCGACATCTGTTGATGAAGAGTCTACAGCCAGATGGTGCARAC	1859	QY	2879	TATACCGCTCGACAGATCTTGAACACTGGGAATTCOCGCTGAATCACCTTTGACCTCA	2938
DB	113785	GAGTTCACATGACACATCGACATCTGTTGATGAAGAGTCTACAGCCAGATGGTGCARAC	113726	DB	112705	TATACCGCTCGACAGATCTTGAACACTGGGAATTCOCGCTGAATCACCTTTGACCTCA	112646
QY	1860	CTGTATCCATPACGAAGCAGCTGAGAGTATGAGCAAGATCGCTGCACAAATCCGTTGA	1919	QY	2939	GTGATGCAACACTGGTCTCTCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT	2998
DB	113725	CTGTATCCATPACGAAGCAGCTGAGAGTATGAGCAAGATCGCTGCACAAATCCGTTGA	113666	DB	112645	GTGATGCAACACTGGTCTCTCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT	112586
QY	1920	GTTCAAACTCTTGGCATCGCGGAAACGG-CACATCGCTTTCAATGAACCATCATCTTCT	1978	QY	2999	CCAACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATTTCT	3058
DB	113665	GTTCAAACTCTTGGCATCGCGGAAACGGCCACATCGCTTTCAATGAAGCATCATCTTCT	113606	DB	112585	CCAACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGCTGCTGATTTCT	112526
QY	1979	CTGTAGGACTGACAAAGTCCAGCGCTGCACCTTAAACTGTGAGGACAACTCTGA	2038	QY	3059	GTCCACAGAGTTCGACCGAATCCAGCATGAGTACTCACTACGCAAGCTCTCACCAGT	3118
DB	113605	CTGTAGGACTGACAAAGTTCAGGCGCTGCACCTTAAACTGTGAGGACAACTCTGA	113546	DB	112525	GTCCACAGAGTTCGACCGAATCCAGCATGAGTACTCACTACGCAAGCTCTCACCAGT	112466
QY	2039	TTCTTCAACACCATGCAAGAGTTCACACCCAGCGCTCACCCAGGGTTTGGGCACTTG	2098	QY	3119	CGGATATGCTCGTGCACAACTTGAAGAACAGCTTCGCGCTTTCGAGGATTCAGCG	3178
DB	113545	TTCTTCAACACCATGCAAGAGTTCACACCCAGCGCTCACCCAGGGTTTGGGCACTTG	113486	DB	112465	CGGATATGCTCGTGGCAAGCTTGAAGAACAGCTTCGCGCTTTCGAGGATTCAGCG	112406
QY	2099	TCCCGCGGCAAAACATCTGTTGTTGGCACTGTTGAGGAAAGCCACGCCATCCGC	2158	QY	3179	AGCTGGATTTCCGCGCATGAATTTACGACCGCAGGTTGAGTAAACGGTCTGATCCT	3238
DB	113485	TCCCGCGGCAAAACATCTGTTGTTGGCACTGTTGAGGAAAGCCACGCCATCCGC	113426	DB	112405	AGCTGGATTTCCGCGCATGAATTTACGACCGCAGGTTGAGTAAACGGTCTGATCCT	112346
QY	2159	GGAACCTGTGGAAGGCCACTGACTGCTTCTTGCCAGGTTCCATCTGTAGATGCACAA	2218	QY	3239	GGCTCGTGGGTGGATGGGCTGCCCGCAGGATGATCACCCCAACAGTTCCACAGGAAG	3298
DB	113425	GGAACCTGTGGAAGGCCACTGACCAGCATGTGCCAGGTTCCATCTGTAGATGCACAA	113366	DB	112345	GGCTCGTGGGTGGATGGGCTGCCCGCAGGATGATCACCCCAACAGTTCCACAGGAAG	112286
QY	2219	ATGCCACCATCATCTGTTGATGAGCAGCAGTATCCAAAGCTGGAACGCTGATCCTAC	2278	QY	3299	GATGGGTGCATGCTGCTGACTGTGCCCGCAGCTTCATTTCCGCAACACCGGATCTACC	3358
DB	113365	ATGCCACCATCATCTGTTGATGAGCAGCAGTATCCAAAGCTGGAACGCTGATCCTAC	113306	DB	112285	GATGGGTGCATGCTGCTGACTGTGCCCGCAGCTTCATTTCCGCAACACCGGATCTATC	112226
QY	2279	CGTCTCATGGAGCAATTAAGCTGCCTAGAACCAAAAGAAAGTACTGTGTGGGGCTA	2338	QY	3359	AAGAGCTCCTTCTCCAGAGGGGAGTCGGGGGTAAATCAGATCTGTATTAGGTTCTGAAC	3418
DB	113305	CGTCTCATGGAGCAATTAAGCTGCCTAGAACCAAAAGAAAGTACTGTGTGGGGCTA	113246	DB	112225	AAGAGCTCCTTCTCCAGAGGGGAGTCAGGGGTAAATCAGATCTGTATTAGGTTCTGAAC	112166
QY	2339	TGCACACAGAACTTCCAGTTTGGCCCTGCGTACCATGTGACTCTCCGAGGGCAGGC	2398	QY	3419	CTGTCCAGATGAGACATCCGAGGCAATATTTCCTTCGAGTGGGATGGTGTCCGTTGCTG	3478
DB	113245	TGCACACAGAACTTCCAGTTTGGCCCTGCGTACCATGTGACTCTCCGAGGGCAGGC	113186	DB	112165	CTGTCCAGATGAGACATCCGAGGCAATATTTCCTTCGAGTGGGATGGTGTCCGTTGCTG	112106
QY	2399	TCAATGATCCCAACGGAATGATGCTGATGGAGATACCCCTCCACGCTCTACTACCAAGCAG	2458	QY	3479	TGGATCTGTGATGCTGCTGCTGAGTAGTGAAGTAAACCTTGGCGAATTAGTGATCCGG	3538
DB	113185	TCAATGATCCCAACGGAATGATGCTGATGGAGATACCCCTCCACGCTCTACTACCAAGCAG	113126	DB	112105	TGGATCTGTGATGCTGCTGCTGAGTAGTGAAGTAAACCTTGGCGAATTAGTGATCCGG	112046
QY	2459	ATCCAGGTTCCCTTCGCAACCAAGCCAGCGGTGGGCTCACACCAACCGCGGTTGA	2518	QY	3539	ACGATAATACAGCATTTAGATTAAGTGCAGGTGATGACAGGTTTCAATTCGCTTTCCGG	3598
DB	113125	ATCCAGGTTCCCTTCGCAACCAAGCCAGCGGTGGGCTCACACCAACCGCGGTTGA	113066	DB	112045	ACGATAATACAGCATTTAGATTAAGTGCAGGTGATGACAGGTTTCAATTCGCTTTCCGG	111987
				QY	3599	GCCTTCAAAGGTGACACTATTGAGAGATAAGTCAATAAAAGGCTCTTTTGTGCGCAATT	3658

Db	111986	GCCTTCAAGGTGACACTATTGAGAGATAAGTCATAAAAGGGCTTTTGGCGAATT	111927	Db	110906	GATTTCCGTGTTCCAGTCGGCGGTCTGCTCTTCGTTCTACTCACCAATCGTCATC	110847
Qy	3659	GTACAAATACCTTCGCAAAATCCCTTGATCGGACACAAATAACAGGTTTAAATATGTTTFA	3718	Qy	4739	ACTGCTCTGCACAGTCTTCCCGCCAAATTAGCTGAGCTGTTTCAACAGGGTGGATCC	4798
Db	111926	GTACAAATACCTTCGCAAAATCCCTTGATCGGACACAAATAACAGGTTTAAATATGTTTFA	111867	Db	110846	ACTGCTCTGCACAGTCTTCCCGCCAAATTAGCTGAGCTGTTTAAACAGGGTGGATCC	110787
Qy	3719	GCCTTTGAACAAACATTCATGCTGAATATTTTGTCTTCCCGGTTTAAAGAGAAATTC	3778	Qy	4799	TTCATCTTCGCAACCGCATCCATGSCCAATATCGCGCAGGGTGCAGATGTTTGGCAGTG	4858
Db	111866	GCCTTTGAACAAACATTCATGCTGAATATTTTGTCTTCCCGGTTTAAAGAGAAATTC	111807	Db	110786	TTCATCTTCGCAACCGCATCTATGGCTAATATCGCCAGGGTGGCGATGTTTGGCAGTG	110727
Qy	3779	ATGGACATAAGGACCTCGCGCAACGCATCTCTGCGGACATTCCTGGCGGCAAGACACAT	3838	Qy	4859	TTCCTTCAGGACAGTGAAGAGCTCAAGGCGCTTCAGAGTGTCTCAGGTGTCTCCGCT	4918
Db	111806	ATGGACATAAGGACCTCGCGCAACGCATCTCTGCGGACATTCCTGGCGGCAAGACACAT	111747	Db	110726	TTCCTTCAGGACAGTGAAGAGCTCAAGGCGCTTCAGAGTGTCTCAGGTGTCTCCGCT	110667
Qy	3839	GTCCCGCGCACACTGTGCAACGCTTTACGCTCTGCTCAAAAGACACCAAGGATG	3898	Qy	4919	GTCTTGTGATTACAGAGCTTCGATCTTCGCTGTGAACCTTCGCCCTCGCTGGCCGCTTC	4978
Db	111746	GTCCCGCGCACACTGTGCAACGCTTTACGCTCTGCTCAAAAGACACCAAGGATG	111687	Db	110666	GTCTTGTGATTACAGAGCTTCGATCTTCGCTGTGAACCTTCGCCCTCGCTGGCCGCTTC	110607
Qy	3899	GATCGGCAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	3958	Qy	4979	TACATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	5038
Db	111686	GATCGGCAAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	111627	Db	110606	TTCATCGGTATCGGTACCGCAGCTATCGGTGCGCTTGTGATGTCACCTCTTAAATATCAAG	110547
Qy	3959	TTCAGATCATCTCGCGGCGAGCGGATGATGATGATGATGATGATGATGATGATGATG	4018	Qy	5039	GCAGTTGCGTTGGGGGCTGCGAGGTTCTTGGGTGTTGTTCTATTGATGCTCCAGATATG	5098
Db	111626	TTCAGATCATCTCGCGGCGAGCGGATGATGATGATGATGATGATGATGATGATGATG	111567	Db	110546	GCAGTTGCGTTGGGGGCTGCGAGGTTCTTGGGTGTTGTTCTATTGATGCTCCAGATATG	110487
Qy	4019	ACCTCCAAAGACATCGCTGTGTCACAGACAGCTCAAGATGTTGTGCTTAACACGCG	4078	Qy	5099	GTATGTTCTTGGTTTGGCGGCTAGTACCTTTGTGTCGATCTCGGCGAGCGATGCT	5158
Db	111566	ACCTCCAAAGACATCGCTGTGTCACAGACAGCTCAAGATGTTGTGCTTAACACGCG	111507	Db	110486	GTATGTTCTTGGTTTGGCGGCTAGTACCTTTGTGTCGATCTCGGCGAGCGATGCT	110427
Qy	4079	AACCTGTTACGCGTCTGTAAGGATTTGGCGGACATTTCTGCTCCGCTGATTCACATC	4138	Qy	5159	TATGCGCTTACTTGGTTGCGCGCAACGCGAGCATGATGATGATGATGATGATGATGATG	5218
Db	111506	AACCTGTTACGCGTCTGTAAGGATTTGGCGGACATTTCTGCTCCGCTGATTCACATC	111447	Db	110426	TATGCGCTTACTTGGTTGCGCGCAACGCGAGCATGATGATGATGATGATGATGATGATG	110367
Qy	4139	TGCTTGGTGGCGGCTGCTCATGCTATCAACAATGTGTTGTTGCGGAGGATGTTTC	4198	Qy	5219	GTGCTGTCAGGAACACCAAGCCGAGCAGAAACCCGAGAAATTTTCAAAAGCATTC	5278
Db	111446	TGCTTGGTGGCGGCTGCTCATGCTATCAACAATGTGTTGTTGCGGAGGATGTTTC	111387	Db	110366	GTGCTGTCAGGAACACCAAGCCGAGCAGAAACCCGAGAAATTTTCAAAAGCATTC	110307
Qy	4199	GGTCCGCAATCACTGTTGGAGATGTTCCCTCAGATCAGCGGTTTCTGATGATGATCAAC	4258	Qy	5279	ACCATCATCCAGGACCTTTGACCGGTGAAGCTATCGACATCGACAGCTCAGCGATGCG	5338
Db	111386	GGTCCGCAATCACTGTTGGAGATGTTCCCTCAGATCAGCGGTTTCTGATGATGATCAAC	111327	Db	110306	ACCATCATCCAGGACCTTTGACCGGTGAAGCTATCGACATCGACAGCTCAGCGATGCG	110247
Qy	4259	CTGATGSCATCTCGCGGTTTCGGCTTCTGCGGAGTGTGTTGTTTCCCGCAACCAAG	4318	Qy	5339	ATGTTTGGCAGCGGAAAGCTTTGGCTCAGGTGTTGCGATGTCGCCACCAAGGGCAGTG	5398
Db	111326	CTGATGSCATCTCGCGGTTTCGGCTTCTGCGGAGTGTGTTGTTTCCCGCAACCAAG	111267	Db	110246	ATGTTTGGCAGCGGAAAGCTTTGGCTCAGGTGTTGCGATGTCGCCACCAAGGGCAGTTA	110187
Qy	4319	CGTTTCGGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	4378	Qy	5399	GTCTTCCAGCAGTACGCGGAAAGATCGTGTGGCTTCCCATCTGTCACGCTTTTGCAGTC	5458
Db	111266	CGTTTCGGTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	111207	Db	110186	GTCTTCCCGGTGAGTGAAGATTTGTTGTCATTCCTCATCTGCGCATGTTTTCGCACTT	110127
Qy	4379	CTGTTTAAACGGCTACGACGTGCGCGCCACCATGACCGCGGCGAAATGCCAATGTGCTC	4438	Qy	5459	CGCATTAAGGCTGAGGATGTTCCCAATGTGATATCTTGTGATGTCACATGTTTTCGACACC	5518
Db	111206	TTGGTGAACGGCTACGACGTGCGCGCCACCATGACCGCGGCGAAATGCCAATGTGCTC	111147	Db	110126	CGCACCAAGCTGAGGATGTTCCCAATGTGATATCTTGTGATGTCACATGTTTTCGACACA	110067
Qy	4439	CTGTTTGGTTTGGATGTTGCTCAAGCTGTTTACAGGCGACCGTGTCTCTGTCGCTG	4498	Qy	5519	GTAAACCTCAACGGCACGCTTTAAACCGGTGAAGAAGAGGGGATGAAGTCAAGCA	5578
Db	111146	CTGTTTGGTTTGGATGTTGCTCAAGCTGTTTACAGGCGACCGTGTCTCTGTCGCTG	111087	Db	110066	GTAAACCTCAACGGCACGCTTTAAACCGGTGAAGAAGAGGGGATGAAGTCAAGCA	110007
Qy	4499	GTCTCTGGATCTGCAACGATCGAGAGTTCCTGCAAGCGACTCATGGCAGTCA	4558	Qy	5579	GGGAGCTGCTGTGTGAATTCGATATTCGATGTCATTAAGGCTGAGGTTTATGAGTAACC	5638
Db	111086	GTCTCTGGATCTGCAACGATCGAGAGTTCCTGCAAGCGACTCATGGCAGTCA	111027	Db	110006	GGGAGCTGCTGTGTGAATTCGATATTCGATGTCATTAAGGCTGAGGTTTATGAGTAACC	109947
Qy	4559	GACTTCCTGATCACCACGATGTTGACTCTGCTCCTCAGCGGCTTCCTTACGTTCAATGCT	4618	Qy	5639	ACGCGGATGTTTTCGAAATTAAGAAACCGGACCTGTAAACACTTACGTTTGGGC	5698
Db	111026	GACTTCCTGATCACCACGATGTTGACTCTGCTCCTCAGCGGCTTCCTTACGTTCAATGCT	110967	Db	109946	ACGCGGATGTTTTCGAAATTAAGAAACCGGACCTGTAAACACTTACGTTTGGGC	109887
Qy	4619	ATTGGTCCACATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	4678	Qy	5699	GAAATTAAGCGGGAGCCAACTGCTCAACGTCGCAAGAAAGAGCGGTCGCGACACA	5758
Db	110966	ATTGGTCCACATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	110907	Db	109886	GAAATTAAGCGGGAGCCAACTGCTCAACGTCGCAAGAAAGAGCGGTCGCGACACA	109827
Qy	4679	GATTTCGGTGGTTCAGTTCGCGGCTGCTGTTTCGGTCTGCTTCTACCTACCAATCGTTATC	4738	Qy	5759	CCATAAGTTGAACCTTTCAGTGTTCGACACAGGTTAGCTAGGAGGCTGACTCTAGC	5818
Db				Db	109826	CCATAAGTTGAACCTTTCAGTGTTCGACACAGGTTAGCTAGGAGGCTGACTCTAGC	109767

QY 5819 ATCTTTGACACCGGTACCGCTCGAGCTTAAACCTGTTCAACCCAGGTGATGCC 5878
 Db 109766 ATCTTTGACACCGGTACCGCTCGAGCTTAAACCTGTTCAACCCAGGTGATGCC 109707
 QY 5879 TCGGTGACTGCTGTGCGCCACCGCGCAATCTTACCCACATGGACATGTTGTTCA 5938
 Db 109706 TCGGTGACTGCTGTGCGCCACCGCGCAATCTTACCCACATGGACATGTTGTTCA 109647
 QY 5939 GCAGTAGCTTGTATTTTGGCGCGCTG 5967
 Db 109646 GCAGTAGCTTGTATTTTGGCGCGCTG 109618

RESULT 3

AAAT34177
 ID AAAT34177 standard; DNA; 6911 BP.
 XX AC AAAT34177;
 XX AC AAAT34177;
 DT 22-OCT-1996 (first entry)
 XX DE Coryneform sucrose gene.
 XX KW Sucrose; coryneform; L-amino acid; L-lysine; L-glutamic acid;
 XX KW nucleic acid; sucrose fermentation; ss.
 XX OS Brevibacterium lactofermentum strain ATCC 13869.
 XX PH Key Location/Qualifiers
 XX CDS 342..1508
 FT /tag= a
 FT /label= ORF-F1
 FT /product= product shows 24% homology to E. coli
 FT N-acetylglucosamine-6-phosphate
 FT deacetylase
 FT CDS 2338..3612
 FT /tag= b
 FT /label= ORF-F2
 FT /product= sucrose
 FT 4438..5358
 FT /tag= c
 FT /label= ORF-F3
 FT /product= product shows 36% homology to B. subtilis
 FT Up-N-acetylmuramoylalanyl-D-glutamyl-
 FT meso-6-diaminopimelate synthetase
 FT CDS 5570..5580
 FT /*tag= d
 FT /label= ORF-F2
 FT /product= product shows 39% homology to B. subtilis
 FT phospho-N-acetylmuramoylpentapeptide
 FT transferase
 XX EP724017-A2.
 XX 31-JUL-1996.
 XX 29-JAN-1996; 96EP-0101228.
 XX 30-JAN-1995; 95JP-0012361.
 XX (AJIN) AJINOMOTO KK.
 XX (AJIN) AJINOMOTO CO INC.
 XX Hiroshi M, Nagase K, Nakamatsu T, Otsuna S, Sugimoto M;
 XX Tsuchiya M, Yasuhiko Y, Matsui H, Yoshihara Y;
 XX WPI; 1996-343532/35.
 XX Sucrose gene from Coryneform bacteria - used to prepare recombinant
 PT microorganisms for improved prodn. of L-amino acids and nucleic
 PT acids by fermentation of sucrose

PS Claim 2; Page 20-24; 24pp; English.
 XX A DNA fragment (AAAT34177) includes an open reading frame (ORF-F2)
 CC coding for sucrose. It was obtd. by screening a genomic library
 CC of Brevibacterium lactofermentum ATCC 13869 with probes (see also
 CC AAAT34178-80) based on homologous regions of known sucrose, levanase
 CC and invertase genes. An isolated fragment was inserted into pSAC4,
 CC giving plasmid pSSM30. Escherichia coli JM109/pSSM30 was named
 CC AJ13047 and deposited as FERM BP-4800. Over-expression of the
 CC sucrose gene in a coryneform bacterium improves the ability of the
 CC host to assimilate sucrose, resulting in increased fermentation
 CC rates and improved productivity of L-amino acids (esp. L-lysine and
 CC L-glutamic acid) and nucleic acids from raw materials contg.
 XX sucrose.
 XX Sequence 6911 BP; 1575 A; 1784 C; 1892 G; 1660 T; 0 other;
 SQ Query Match 61.8%; Score 3687; DB 17; Length 6911;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTCGTCGACGCCACCATTCATGCTGGTGCACCGAGCTTGGAGGCTTTCTACATCT 60
 Db 1 AGTCGTCGACGCCACCATTCATGCTGGTGCACCGAGCTTGGAGGCTTTCTACATCT 60
 QY 61 ACCTCCCGTCGCGTGGAGTGGGTTCATTACGGGTGGGATCACGCCGGTGAAGTGGCG 120
 Db 61 ACCTCCCGTCGCGTGGAGTGGGTTCATTACGGGTGGGATCACGCCGGTGAAGTGGCG 120
 QY 121 GAACCCATGGTTCCTTGTGGTGGAGGACAGTCCGGTGCAGAGTTCCTCAAGTG 180
 Db 121 GAACCCATGGTTCCTTGTGGTGGAGGACAGTCCGGTGCAGAGTTCCTCAAGTG 180
 QY 181 TCTGCAGTTTTAAAGTATGATCATCATCAGCTTGAAGGCTGAGTAATTCAGTAGAGCTG 240
 Db 181 TCTGCAGTTTTAAAGTATGATCATCATCAGCTTGAAGGCTGAGTAATTCAGTAGAGCTG 240
 QY 241 CAACAGCAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAGACATCATA 300
 Db 241 CAACAGCAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAGACATCATA 300
 QY 301 CGTCTATGCTTGTGGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGATT 360
 Db 301 CGTCTATGCTTGTGGAAGAACCAATAACCTCAGAAAGATGGCAGAGTGGTGATT 360
 QY 361 ATCAGAAATGCAGGTCAAGTCAAGTAAATAAATGAGGGAAGATTTCCCGCCCTCG 420
 Db 361 ATCAGAAATGCAGGTCAAGTCAAGTAAATAAATGAGGGAAGATTTCCCGCCCTCG 420
 QY 421 GGGTGATTGATGGCTTTCTCAACTCGAAACGGCATCATCACGGAACCTCTCTGGAAC 480
 Db 421 GGGTGATTGATGGCTTTCTCAACTCGAAACGGCATCATCACGGAACCTCTCTGGAAC 480
 QY 481 CAGCACCTAAAACGCGAGTTCACCCGAGTCCCGGCGGTTTCCACGAGCAGCAGCGGA 600
 Db 481 CAGCACCTAAAACGCGAGTTCACCCGAGTCCCGGCGGTTTCCACGAGCAGCAGCGGA 600
 QY 541 ATCTTCATAATCACGGTGGAAAGCGTGGCGGTTTCCCTACGGGAACGAGCAGCGGA 660
 Db 541 ATCTTCATAATCACGGTGGAAAGCGTGGCGGTTTCCCTACGGGAACGAGCAGCGGA 660
 QY 601 GGAACACCGCGAGTATCACCGCGNACATGGCAGCAGCGTGTGTCGACAGCATGGTTT 660
 Db 601 GGAACACCGCGAGTATCACCGCGNACATGGCAGCAGCGTGTGTCGACAGCATGGTTT 660
 QY 661 CGGCGCGCGCTGACCACTGGCAGCGGAGTGGAAACCTTATCCCTTGTGTGAAGAGG 720
 Db 661 CGGCGCGCGCTGACCACTGGCAGCGGAGTGGAAACCTTATCCCTTGTGTGAAGAGG 720
 QY 721 TCCTGCTGTGGGATTCACCTCGAGGCGCCCTTTTCATCAACGCATCGCGTGTGGTCTC 780
 Db 721 TCCTGCTGTGGGATTCACCTCGAGGCGCCCTTTTCATCAACGCATCGCGTGTGGTCTC 780

QY 781 AAAACCGGATTTCAATTTTCCCGCAACCAACAGATCTGCCGGGTGATCCATCGG 840
DB 781 AAAACCGGATTTCAATTTTCCCGCAACCAACAGATCTGCCGGGTGATCCATCGG 840
QY 841 GAAAGGTTGGATCAATGATCACATAGTACGCGGAAACTGACAATCTTCTGAGCTTC 900
DB 841 GAAAGGTTGGATCAATGATCACATAGTACGCGGAAACTGACAATCTTCTGAGCTTC 900
QY 901 TCGATCTCTCGGAGCGCACCATCATCTCTTCCCTCGGGCACACTGATGCAAGATTG 960
DB 901 TCGATCTCTCGGAGCGCACCATCATCTCTTCCCTCGGGCACACTGATGCAAGATTG 960
QY 961 ATACCACCTACAGCGCAATTCCTTGGCTAAAGAGAAAAATGAGCGGTACGCGTACGC 1020
DB 961 ATACCACCTACAGCGCAATTCCTTGGCTAAAGAGAAAAATGAGCGGTACGCGTACGC 1020
QY 1021 ATTGTTCATAGGATGCCCTCCGCTGATCATAGGCGTCCCGGAGCGGTGGCGCTTGC 1080
DB 1021 ATTGTTCATAGGATGCCCTCCGCTGATCATAGGCGTCCCGGAGCGGTGGCGCTTGC 1080
QY 1081 TTGCTGGGACGTCGCGGAGCGCATATGTTGAGTTGATCGCGGAGCGGTGCATTTGG 1140
DB 1081 TTGCTGGGACGTCGCGGAGCGCATATGTTGAGTTGATCGCGGAGCGGTGCATTTGG 1140
QY 1141 CCGATGAAAGGTCGATCTAGCTCTTCCCAACAAGCGCTTTTTCATCACGAGCGCATGG 1200
DB 1141 CCGATGAAAGGTCGATCTAGCTCTTCCCAACAAGCGCTTTTTCATCACGAGCGCATGG 1200
QY 1201 AAGCGCGGAAATGCCAGCGGTAGTACATTTTGGCGTTTGAAGCTCACCGTACCG 1260
DB 1201 AAGCGCGGAAATGCCAGCGGTAGTACATTTTGGCGTTTGAAGCTCACCGTACCG 1260
QY 1261 ATGAGTGCAGCTGCGGAGTGGCGGCGCATCGCGGGGACACGACACACTAGCGA 1320
DB 1261 ATGAGTGCAGCTGCGGAGTGGCGGCGCATCGCGGGGACACGACACACTAGCGA 1320
QY 1321 GTCAGTTCTGTCACCGTGCAGAGGATGATGAGTTATCGAGCGGACCTCCACACT 1380
DB 1321 GTCAGTTCTGTCACCGTGCAGAGGATGATGAGTTATCGAGCGGACCTCCACACT 1380
QY 1381 CAACCGTGCAGTAAATCTCGCTTGGCGATCAGAAATCGCTAAATCCAAACCTG 1440
DB 1381 CAACCGTGCAGTAAATCTCGCTTGGCGATCAGAAATCGCTAAATCCAAACCTG 1440
QY 1441 CAAATTTTGTGCTTTGACTCAACCGCGAGTGCAGGTTGAGTTCATCAAG 1500
DB 1441 CAAATTTTGTGCTTTGACTCAACCGCGAGTGCAGGTTGAGTTCATCAAG 1500
QY 1501 TACTTTAAGTACGATTAATCTCCTGATTTAAAGGAGTCCGACCATGGAATCACTA 1560
DB 1501 TACTTTAAGTACGATTAATCTCCTGATTTAAAGGAGTCCGACCATGGAATCACTA 1560
QY 1561 TCTCAAAGAGGAGCAAGAGTGGCAAGCAGTTGAGTTCCTAATGCGACCTTCGCCA 1620
DB 1561 TCTCAAAGAGGAGCAAGAGTGGCAAGCAGTTGAGTTCCTAATGCGACCTTCGCCA 1620
QY 1621 ACAAGGTGGAACCTTGGGCTTGCACAGAGTCCCTCACCACTGAGTACCTACAGAGC 1680
DB 1621 ACAAGGTGGAACCTTGGGCTTGCACAGAGTCCCTCACCACTGAGTACCTACAGAGC 1680
QY 1681 TCATTCGATGATGAGCTGGGAGTGTCAATCAAGAACTGCAAGGCAATCTTGTGG 1740
DB 1681 TCATTCGATGATGAGCTGGGAGTGTCAATCAAGAACTGCAAGGCAATCTTGTGG 1740
QY 1741 ATGAATACGTGGGACTAACCGGTGACGATGAACAGCTACTTTAAACCATTCGCAAG 1800
DB 1741 ATGAATACGTGGGACTAACCGGTGACGATGAACAGCTACTTTAAACCATTCGCAAG 1800
QY 1801 AGTTCACCTGACACATCGATCTTGTATGAAGAGTCTACAGCCAGATGGTGAAC 1860
DB 1801 AGTTCACCTGACACATCGATCTTGTATGAAGAGTCTACAGCCAGATGGTGAAC 1860
QY 1861 CTGATCCATACGAAGAGCTGACAGATGATGAGCAAAAGATCGCTGCAAAATCCGTTGAG 1920

DB 1861 CTGATCCATACGAACAGCTGCAGAGTATGAGGCAAGATCGCTCAGAATCCGTTGAG 1920
QY 1921 TTCAAATCCTTGGCATCGGCGGAAAGCGCATCGCTTTCATTGAACCATCATCTTCCT 1980
DB 1921 TTCAAATCCTTGGCATCGGCGGAAAGCGCATCGCTTTCATTGAACCATCATCTTCCT 1980
QY 1981 GTCAGACTGACAAAGGTCCAGCGCTGCACCCCTAAAACCTGTGGAGGACACGCTCGATT 2040
DB 1981 GTCAGACTGACAAAGGTCCAGCGCTGCACCCCTAAAACCTGTGGAGGACACGCTCGATT 2040
QY 2041 CTTCAACACCTACGAGAGGTCCCAACCCAGCGCTCACCCAGGTTTGGGCACTTTGTC 2100
DB 2041 CTTCAACACCTACGAGAGGTCCCAACCCAGCGCTCACCCAGGTTTGGGCACTTTGTC 2100
QY 2101 CCGCGCGCAAAACATCGTGTGGTGGCAACTGTTGAAGGAAAAAGCGGACGATCCGGG 2160
DB 2101 CCGCGCGCAAAACATCGTGTGGTGGCAACTGTTGAAGGAAAAAGCGGACGATCCGGG 2160
QY 2161 AACTGTGGAAGCGCCAGTACCTCTTGTGCCAGGTTCCATCTCTGTAGATGCACAACT 2220
DB 2161 AACTGTGGAAGCGCCAGTACCTCTTGTGCCAGGTTCCATCTCTGTAGATGCACAACT 2220
QY 2221 GGCACCATCATCGTTGGATGAAGCAGCAGTATCCAAAGCTGGAACCGCTGATCACTACCG 2280
DB 2221 GGCACCATCATCGTTGGATGAAGCAGCAGTATCCAAAGCTGGAACCGCTGATCACTACCG 2280
QY 2281 TCTCATGAGCAATTAAGCTGCGCTAGNAACAAAGAAAGTACTGTGTGGGCTATG 2340
DB 2281 TCTCATGAGCAATTAAGCTGCGCTAGNAACAAAGAAAGTACTGTGTGGGCTATG 2340
QY 2341 CACACAGAACTTTCCAGTTTGGCGCTGATGACCTTCCCGCAGGCGAGGCTC 2400
DB 2341 CACACAGAACTTTCCAGTTTGGCGCTGATGACCTTCCCGCAGGCGAGGCTC 2400
QY 2401 AATGATCCCAAGCAATGATGATGAGATACCTCCAGCTTACTTACAGCAGCAT 2460
DB 2401 AATGATCCCAAGCAATGATGATGAGATACCTCCAGCTTACTTACAGCAGCAT 2460
QY 2461 CAGGTTTCCCTTCCACAAAGCGACCGCTGGGCTCACACACACGCGGTTGACC 2520
DB 2461 CAGGTTTCCCTTCCACAAAGCGACCGCTGGGCTCACACACACGCGGTTGACC 2520
QY 2521 GAGCCGACGATGATGAGTGGACGACCTTGCAGGCTTTCACCGGATGCATCCAT 2580
DB 2521 GAGCCGACGATGATGAGTGGACGACCTTGCAGGCTTTCACCGGATGCATCCAT 2580
QY 2581 GACCTGATGATGATGATGATGAGGCTTACTTACTGAGGCGACACTTAAACTTTC 2640
DB 2581 GACCTGATGATGATGATGATGAGGCTTACTTACTGAGGCGACACTTAAACTTTC 2640
QY 2641 TACACGGCAACCTAAAATTAAGGAAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2700
DB 2641 TACACGGCAACCTAAAATTAAGGAAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2700
QY 2701 GAGGACCAACTGCGGTGATGCGGCTTTCATCGCTTCCCTTAAATTCGCTTATC 2760
DB 2701 GAGGACCAACTGCGGTGATGCGGCTTTCATCGCTTCCCTTAAATTCGCTTATC 2760
QY 2761 GAGGACCAACTGCGGTGATGCGGCTTTCATCGCTTCCCTTAAATTCGCTTATC 2820
DB 2761 GAGGACCAACTGCGGTGATGCGGCTTTCATCGCTTCCCTTAAATTCGCTTATC 2820
QY 2821 GATGTTGGAACATGTTCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2880
DB 2821 GATGTTGGAACATGTTCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2880
QY 2881 TACCGCTGACAGATCTTGAAGGAAATTCCTCGGTTGAATCAATCAGCTTTGACCTCAGT 2940
DB 2881 TACCGCTGACAGATCTTGAAGGAAATTCCTCGGTTGAATCAATCAGCTTTGACCTCAGT 2940
QY 2941 GATGCAACCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3000

Query Match	24.4%	Score 1457	DB 22	Length 1656
Best Local Similarity	93.2%	Pred. No. 0		
Matches 1545	Conservative	0	Mismatches 110	Indels 2
Gaps	2			
QY 2031	ACGTCGTATCTTCAACACCATCGAAGAGTCCCAACCCACGCGTCACCCACGGTTTGG	2090		
DB 1	ACGCCGATCTTCAACACTATCGAAGAGTCCCAACCCACGCGTTGACCCAGGGCTTGG	60		
QY 2091	GCATTTGTCCCGCGCAAAACATCGTTGTGGCACTGTGTGAAGGAAAAAGCCGAC	2149		
DB 61	GTACTTTGTCCCGCGCAAAATATCGTTGTGTGGCACTGTGGCAAGGAAAAAGCAGCA	120		
QY 2150	GCCATCCGCGGAACGTGTGAAGGCCACGTGACTGCTTCTTGGCCAGAGTCCATCTCTAG	2209		
DB 121	GCCATCCGCGGAACGTGTGAAGGTCCAGTCACTGCTTCTTGGCCAGAGTTCATCTCTCAA	180		
QY 2210	ATGCACAACATGCCACCATCATCGTTGGATGAAGCAGCAGTATCCAAAGCTGGAAAAAGCT	2269		
DB 181	ATGCACAACAGCCACCATCATCGTTGATGAAGCAGCAGCATCCAAAGCTGAAAAATGCT	240		
QY 2270	GATCACTACCGTCTCATGGAGCAATTAAGCTGGCTAGAAACAAAAAGAAAGTACTGT	2329		
DB 241	GACCATTACCGTCTCATGGAGCAATTAAGCTGGCTAGAAACAAAAAGAAAGTACTGT	300		
QY 2330	GTGGGGTATGCACAGAACTTTCCAGTTTGGCCCTGCGTACCATGTGACTCTCCGC	2389		
DB 301	GTGGGGTATGCACAGAACTTTCCAGTTTGGCCCTGCGTACCATGTGACTCTCCGC	360		
QY 2390	AGGCAGGCTCAATGATCCCAACGGAATGTACGTGATGAGATACCTCCACGTCTACT	2449		
DB 361	AGGCAGACTCAATGATCCCAATGGAATGTACGTGATGAGATACCTCCACGTCTACT	420		
QY 2450	ACAGCAGGATCAGGTTTCCCTTCCACCAAGACGCACCGGTGGGCTCACACACCA	2509		
DB 421	ACCAGCAGGATCAGGTTTCCCTTCCACCAAGACGCACCGGTGGGCTCACACACCA	480		
QY 2510	CGCGTTGACCGGACCGCAGGATTCAGTGGAGCACCTGCCGAGCGCTTTTACCCGG	2569		
DB 481	CGCGTTGACCGGACCGCAGGATTCAGTGGAGCACCTGCCGATGCTCTTTACCCGG	540		
QY 2570	ATGATCCTATGACCTGGATGATCTATTCGGTGGAGCGGATTTACTGACGGCACAC	2629		
DB 541	ATGATCCTATGACCTGGATGATCTATTCGGGCGGAGCGGATTTTCTGACGGCACAC	600		
QY 2630	TTAAACTTTTCTACACCGCAACTTAAAAATTTACGGAAGCGCGCGCCACCAAAACC	2689		
DB 601	TTAAACTTTTCTACACCGCAACCGAAAAATTTACGCGCAAGCGCGCCACCAAAACC	660		
QY 2690	TTGTGGAAGTCGAGGACCAACTTGGGCTGATGGCGGCATTCATCGCGGTTCGGCTAAAA	2749		
DB 661	TCGTGGAAGTCGAGGACCAACTTGGGCTGATGGCGGCATTCATCGCGGTTCGGCTAAAA	720		
QY 2750	ATCCGCTTATCGAGGACCGCCAGCGGTTTCACACCCCATTCACCGGATCCCATGATCA	2809		
DB 721	ATCCGCTTATCGAGGACCGCCAGCGGTTTACGCCCCACTACCGGATCCCATGATCA	780		
QY 2810	GCCTGATGCTGATGTTGNAATGTTCTTGGGGCCCAACGCGAAAACTCACCGGTG	2869		
DB 781	GCCTGATGCTGATGTTGNAATGTTCTTGGGGCTCAGCGGAAAACTCACCGGTG	840		
QY 2870	CACGGCTTCATACCGCTCGACAGATCTTGAAGAACTGGGAATTTCTCCGGTGAATACCT	2929		

XX 27-JUN-2000; 2000WO-1B00973.
XX 01-JUL-1999; 99US-0142691.
XX 23-AUG-1999; 99US-0150310.
XX 03-SEP-1999; 99DE-1042095.
XX 03-SEP-1999; 99DE-1042097.
XX (BADI) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;
XX WPI: 2001-080989/09.
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
XX sugar phosphotransferase system proteins or their portions, useful for
XX typing or identifying C. glutamicum or related bacteria, and as markers
XX for transformation -
XX Claim 3; Page 98-101; 144pp; English.
XX The present invention relates to Corynebacterium glutamicum
XX phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
XX The PTS nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria, the typing or identification of C. glutamicum or
XX related bacteria, as reference points for mapping C. glutamicum genome,
XX and as markers for transformation.
XX SQ Sequence 1527 BP; 304 A; 392 C; 430 G; 401 T; 0 other;
Query Match 23.78; Score 1412.6; DB 22; Length 1527;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1462; Conservative 0; Mismatches 64; Indels 1; Gaps 1;
QY 4259 CTGATGCACTGCGCGGTTCGGTCTTCCAGTGTGGTTCACCGCAACAAAG 4318
DB 1 CTCATGCACTGCGCGGTTCGGTCTTCCAGTGTGGTTCACCGCAACAAAG 60
QY 4319 CGTTTCGGTGGCAATGATTCCTGGGCGCG -CGATTGGTATGCGATGTGTTCGCAAC 4377
DB 61 CGTTTCGGGCAATGATTCCTGGGCGCGCGTATGATGCGATGTGTTCGCGAG 120
QY 4378 CCTGGTTAAGCGCTACGACGTGGCGCCACCATGACCGCGGGGGAATGCAATGTGGTC 4437
DB 121 CTTGGTGAACGGCTACGACGTGGCGCCACCATGGCTGCGGCGGAATGCAATGTGGTC 180
QY 4438 CCTGTTGGTTGATGTGCTCAAGCTGTTACGAGGCAACCGTGTCTTCTGTGCTGTGT 4497
DB 181 CCTGTTGGTTGATGTGCTCAAGCGGTTCACGAGGCAACCGTGTCTTCTGTGCTGTGT 240
QY 4498 GGTCTCTGGATTCTGGCAACGATCGAGAAGTTCCTGCACAAGCGACTCATGGCACTGC 4557
DB 241 GGTTCCTGGATTCTGSCACGATCGAGAAGTTCCTGCACAAGCGACTCAAGGCACTGC 300
QY 4558 AGACTTCCTGATACCCAGTGTGATCTGTGCTGCTCACCGGTTCCTTACGTTCAATGTC 4617
DB 301 AGACTTCCTGATCACTCCAGTGTGATGCTGCTGCTCACCGGATTCCTTACATCATCGC 360
QY 4618 TATTGGTCCCAATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 4677
DB 361 CATTTGGCCCAATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 420
QY 4678 TGATTCGGTGGTCCAGTGGCGGTCTGCTTTTGGTGTGGTGTGCTTACCAATCGGTAT 4737
DB 421 TGATTCGGTGGTCCAGTGGCGGTCTGCTTCTGGTGTGGTGTGCTTACCAATCGGTAT 480
QY 4738 CACTGGTCTGACAGTCCCTTCCCGCAATGTAGCTGGAGCTGTTTAAACAGGTTGGATC 4797
DB 481 CACTGGTCTGACAGTCCCTTCCCGCAATGTAGCTGGAGCTGTTTAAACAGGTTGGATC 540
QY 4798 CTTTCATCTTCGCAACCGCATCCATGGCCCAATATCGCGAGGTCAGCATGTTTGGCAGT 4857

RESULT 7
ABS65346
ID - ABS65346 standard; DNA; 1527 BP.
XX
AC ABS65346;

DB 541 CTTTCATCTTCGCAACGGCATCTATGCTAATAATCCCCAGGTCGGCATGTTTGGCAGT 600
QY 4858 GTTCTTCTCTAGCAAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTTCAGGTGTCTCCGC 4917
DB 601 GTTCTTCTCTAGCAAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTTCAGGTGTCTCCGC 660
QY 4918 TGTTCCTGTTATTCAGAGCCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGCGTCCGCTT 4977
DB 661 TGTTCCTGTTATTCAGAGCCTCGCATCTTCGGTGTGAACCTTCGCTCGCTGCGTCCGCTT 720
QY 4978 CTACATTTGTTATTCAGAGCCTTCGCTGCGCTTTCGATTCGACCTTCCTTATATCAAA 5037
DB 721 CTTTCATGTTTCTTCGCTACCGCATCGTTCGGCTTTCGATTCGACCTTCCTTATATCAAA 780
QY 5038 GCGAGTTCGTTGCGGCTGCGAGTTCCTTCGGTGTGTTCTTCTATGATGCTCCAGATAT 5097
DB 781 GCGAGTTCGTTGCGGCTGCGAGTTCCTTCGGTGTGTTCTTCTATGATGCTCCAGATAT 840
QY 5098 GGTTCATGTTCTTGGTTCGCGGTAGTACCTTTCATCGCATTCGCGCGCAGCGATTCG 5157
DB 841 GGTTCATGTTCTTGGTGTGTCAGTTCCTTTCATCGCATTCGCGCGCAGCGATTCG 900
QY 5158 TTATGGCTTCTTACTTGGTTCGCGCAACGGCAGCATTCGATCCAGATGCAACCGTGCCTCC 5217
DB 901 TTATGGCTTCTTACTTGGTTCGCGCAACGGCAGCATTCGATCCAGATGCAACCGTGCCTCC 960
QY 5218 AGTGGCTGCGAGCAACCAAGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 5277
DB 961 AGTGGCTGCGAGCAACCAAGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1020
QY 5278 CACCATATCCAGGCAACCTTCACGGTGAAGCTATCCACATGACGACGCGTCAGCGATGC 5337
DB 1021 CACCATATCCAGGCAACCTTCACGGTGAAGCTATTCGCGTGAAGCTATTCGCGTGAAGCT 1080
QY 5338 CATGTTTGGCAGCGAAAGCTTGGTTCAGGTTCGCGATCGTCCCGCACCAAGGGGCACT 5397
DB 1081 CATGTTTGGCAGCGAAAGCTTGGTTCGCGCTTCGCGATCGTCCCGCACCAAGGGGCACT 1140
QY 5398 GGTTCACAGTACGCGGAAGATCGTGGTGGCTTCCCATCTGCTCAGCTTCGCGCTTCGCACT 5457
DB 1141 AGTTCCTCGGTGAGTGGAAAGATTTGGTGGCATTCGCGCATCTGCGCATGCTTTCGCGAGT 1200
QY 5458 CCGCACTAAGGCTGAGGATGTTTCCAAATGTGATATCTTGTATGATGATGATGATGATGAT 5517
DB 1201 TCGCACAAGGCTGAGGATGTTTCCAAATGTGATATCTTGTATGATGATGATGATGATGAT 1260
QY 5518 CGTAAACCTCAACGGCACCGACTTTAACCCGCTGAAGAGCAGGCGATGATGATGATGATGATGAT 5577
DB 1261 AGTAAACCTCAACGGCACCGACTTTAACCCGCTGAAGAGCAGGCGATGATGATGATGATGATGAT 1320
QY 5578 AGGGAGCTCTGCTGTGAATTCGATATTGATGCCATTAAAGGCTGAGGCTTATGAGGTAAAC 5637
DB 1321 AGGGAGCTCTGCTGTGAATTCGATATTGATGCCATTAAAGGCTGAGGCTTATGAGGTAAAC 1380
QY 5638 CACGCGGATGTTGTTTCGAATTCAGAAACCGGACCTGTAACACTTACGCTTACGCTTGGG 5697
DB 1381 CACGCGGATGTTGTTTCGAATTCAGAAACCGGACCTGTAACACTTACGCTTACGCTTGGG 1440
QY 5698 CGAAATGAAGCGGGGACCAACCTGCTCAACGTCGCAAGAAAGAGAGCGGTGCCAGCAAC 5757
DB 1441 CGAAATGAAGCGGGGACCAACCTGCTCAACGTCGCAAGAAAGAGAGCGGTGCCAGCAAC 1500
QY 5758 ACCATAGTGAACCTTGAAGTTCG 5784
DB 1501 ACCATAGTGAACCTTGAAGTTCG 1527


```
QY 2778 TTTCACACCCATTACCGGATCCCATGATCAGCCCTGATGGTGTGGAACATGCT 2837
Db 541 TTTCACACCCATTACCGGATCCCATGATCAGCCCTGATGGTGTGGAACATGCT 600
QY 2838 TCTTGGGGCCCAACGCGAAACCTACCGGTGCGAGGTTCTATACCGCTCGACAGATCT 2897
Db 601 TCTTGGGGCCCAACGCGAAACCTACCGGTGCGAGGTTCTATACCGCTCGACAGATCT 660
QY 2898 TGAACACTGGGAATTCCTCGGTGAAATCACCCTTGACCTCAGTGATGACACACCTGGTTC 2957
Db 661 TGAACACTGGGAATTCCTCGGTGAAATCACCCTTGACCTCAGTGATGACACACCTGGTTC 720
QY 2958 TGTCTCTGATCTCTCCCGATGGCTACATGTGGGAATGCCCAACCTTTTACGCTTGC 3017
Db 721 TGTCTCTGATCTCTCCCGTGGCTACATGTGGGAATGCCCAACCTTTTACGCTTGC 780
QY 3018 CGATGAAGAACTGGCGAAGATCTCGACGTGCTGATTTCTGTCCACAGGATTTGACCG 3077
Db 781 CGATGAAGAACTGGCGAAGATCTCGACGTGCTGATTTCTGTCCACAGGATTTGACCG 840
QY 3078 AATCCACGATGAGTTACTCAGTACGCAAGCTCTGACAGTGGGATATGCTGTCACAA 3137
Db 841 AATCCACGATGAGTTACTCAGTACGCAAGCTCTGACAGTGGGATATGCTGTCGCA 900
QY 3138 GCTTGAAGAAACGACTTCGCGTCTTGGAGATTCAGGAGCTGGATTTGGGCATGA 3197
Db 901 GCTTGAAGAAACGACTTCGCGTCTTGGAGATTCAGGAGCTGGATTTGGGCATGA 960
QY 3198 ATTCTAGCCACCGAGTTTCAGTAAACGTTCTGATGCTGCTGCTGGGTGGATGGG 3257
Db 961 ATTCTAGCCACCGAGTTTCAGTAAACGTTCTGATGCTGCTGCTGGGTGGATGGG 1020
QY 3258 GCTGCCGCGCAGATGATCACCACAGTTCGACAGGAGGATGGTGCATGCTGCTGAC 3317
Db 1021 GCTGCCGCGCAGATGATCACCACAGTTCGACAGGAGGATGGTGCATGCTGCTGAC 1080
QY 3318 TGTGCCCGCAAGCTTCATTTGGCGCAACACGAGTCTACCAAGAGCTCTCTTCCACGA 3377
Db 1081 TGTGCCCGCAAGCTTCATTTGGCGCAACACGAGTCTACCAAGAGCTCTCTTCCACGA 1140
QY 3378 GGGGAGTCGGGGTATCAGATCTGTATAGTTCTGAACCTGTCCGAGTAGACATCG 3437
Db 1141 GGGGAGTCAGGGGTATCAGATCTGTATAGTTCTGAACCTGTCCGAGTAGACATCG 1200
QY 3438 AGCAATATTTCCCTCGAGTGGATGGTGTCCGTTTGTGTGATCGTGGTGTGATCG 3497
Db 1201 AGCAATATTTCCCTCGAGTGGATGGTGTCCGTTTGTGTGATCGTGGTGTGATCG 1260
QY 3498 TCGGTAGCTCAGGTAAACCTGGCAATAGTATGATCGCGGAGGATTAATACAGCCATGA 3557
Db 1261 TCGGTAGCTCAGGTAAACCTGGCAATAGTATGATCGCGGAGGATTAATACAGCCATGA 1320
QY 3558 GATAACTGCAGGTGATGGACAG 3579
Db 1321 GATAACTGCAGGTGATGGACAG 1342
```

RESULT 9

AAH67870

ID AAH67870 standard; DNA; 1299 BP.

XX AC

XX AC

XX AC

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 2905.

XX Corynebacterium;

XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX

XX

XX

XX

XX

PN

XX

PD

XX

PF

XX

PR

XX

PR

XX

PR

XX

PR

XX

PA

XX

PI

XX

PI

XX

DR

XX

DR

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

QY

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI: 2001-376931/40.

P-PSDB; AAG92651.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Claim 8; SEQ ID NO: 2905; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 1299 BP; 287 A; 363 C; 350 G; 299 T; 0 other;

Query Match 21.2%; Score 1267.8; DB 22; Length 1299;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1286; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 2327 TGTGTGGGGCTATGCACACAGAACTTCCAGTTTGGGCCCTGCTACCATGTGACTCCTC 2386
Db 2 TGTGTGGGGCTATGCACACAGAACTTCCAGTTTGGGCCCTGCTACCATGTGACTCCTC 61

QY 2387 CGCAGGGCAGGCTCAATGATCCCAACGGAATGTACGTGATGGAGATACCTCCACGCT 2446
Db 62 CGCAGGGCAGGCTCAATGATCCCAACGGAATGTACGTGATGGAGATACCTCCACGCT 121

QY 2447 ACTACACAGCATCCAGGTTTCCCTTCGCACCAAGCAGCGGTGGGCTCACACCA 2506
Db 122 ACTACACAGCATCCAGGTTTCCCTTCGCACCAAGCAGCGGTGGGCTCACACCA 181

QY 2507 CCACGCCGTTGACCGGACCGCAGCGATTTGAGTGAGTGCACCTGCCCGACGCTTTACC 2566
Db 182 CCACGCCGTTGACCGGACCGCAGCGATTTGAGTGAGTGCACCTGCCCGACGCTTTACC 241

QY 2567 CGGATGCATCTTATGACCTGGATGGATCTTCCGTTGGAGCGGTATTTACTGACGSCA 2626
Db 242 CGGATGCATCTTATGACCTGGATGGATCTTCCGTTGGAGCGGTATTTACTGACGSCA 301

QY 2627 CACTTAAACTTTTCTACACCGGCAACCTTAAATTTAGCGGAAAGCGCGGCCACCCCAA 2686
Db 302 CACTTAAACTTTTCTACACCGGCAACCTTAAATTTAGCGGAAAGCGCGGCCACCCCAA 361

QY 2687 ACCTTGTGAGTGCAGGACCACTGGGCTGATGGCGGATTCATCGCGGTTCGCCTA 2746
Db 362 ACCTTGTGAGTGCAGGACCACTGGGCTGATGGCGGATTCATCGCGGTTCGCCTA 421

CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. App, NADPH).

XX
 SQ Sequence 1287 BP; 325 A; 349 C; 326 G; 287 T; 0 Other;

Query Match 20.7%; Score 1237.4; DB 22; Length 1287;
 Best Local Similarity 97.6%; Pred. No. 0;
 Matches 1256; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 242 AACAGAGCCCTCAAGTCGGAAGATAATTAACTAGATCCGTAGACATAAGACATCATAC 301
 DB 1 RACAGAGCCCTCAAGTCGGAAGATAATTAACTAGATCCGTAGACATAAGACATCATAC 60
 QY 302 GTCCTATGCTTCTGGAGGAACCAATAAATCTCAGAAAGATGGCAGAGTTCGATTA 361
 DB 61 GTCCTATGCTTCTGGAGGAACCAATAAATCTCAGAAAGATGGCAGAGTTCGATTA 120
 QY 362 TCAAGAAAATGAGGTCAAGCAGTTTAAAAAATTCAGGGAAGAATTGTTCCCCCTCGG 421
 DB 121 TCAAGAAAATGAGGTCAAGCAGTTTAAAAAATTCAGGGAAGAATTGTTCCCCCTCGG 180
 QY 422 GGTGATGATGCTTCTCCAACTCGAAACGGCATCATCAGGAAGTCTCTGGAGAACC 481
 DB 181 GGTGATGATGCTTCTCCAACTCGAAACGGCATCATCAGGAAGTCTCTGGAGAACC 240
 QY 482 AGCACCTAAAACGAGGATTCACCCGGAATCCACCCGAACTCCCGGATGTTCCCGTTTATTGA 541
 DB 241 AGCACCTAAAACGAGGATTCACCCGGAATTCACCCGGAATTCACCCGGAATGTTTATTGA 300
 QY 542 TTTTCAATATCAGGTGGAACCGTGGCGGTTTCTACGGGAACGACGACCGAGG 601
 DB 301 TTTTCAATATCAGGTGGAACCGTGGCGGTTTCTACGGGAACGACGACCGAGG 360
 QY 602 GAACACCGCGAGTATCACCGGAACATGGCAGACCGTGTGTTCCCAAGCATGTTTC 661
 DB 361 GAATCGCGCGAGTATCACCGGAACATGGCAGACCGTGTGTTCCCAAGCATGTTTC 420
 QY 662 GCGCGCGGTGACGACTGGCAGCGAGTGGAAACCTTATTCCTTGTGTGAAGAGT 721
 DB 421 GCGCGCGGTGACGACTGGCAGCGAGTGGAAACCTTATTCCTTGTGTGAAGAGG 480
 QY 722 CTGTGTCGCGCATTCACCTCGAGGCGCTTTCATCAGCATGCGGTGTGTGTCATCA 781
 DB 481 CTGTGTCGCGCATTCACCTCGAGGCGCTTTCATCAGCATGCGGTGTGTGTCATCA 540
 QY 782 AAACCGGATTTTATTTTCCCGGCAACCCACAGATCTTGCCCGGATGATCCATGCGGG 841
 DB 541 AAACCGGATTTTATTTTCCCGGCAACCCACAGATCTTGCCCGGATGATCCATGCGGG 600
 QY 842 AAAGGTTGGATCAATTCATCAGTACGCGCGGGAACACTGACATCTTCTGAGCTTCT 901
 DB 601 AAAGGTTGGATCAATTCATCAGTACGCGCGGGAACACTGACATCTTCTGAGCTTCT 660
 QY 902 CGATCTCTGCGAGCGCACCATCATCTGCTTCCCTTGGGCACTGATGACAGATTTGA 961
 DB 661 CGATCTCTGCGAGCGCACCATCATCTGCTTCCCTTGGGCACTGATGACAGATTTGA 720
 QY 962 TACCACTACAGGCAATTCCTTGGCTTAAGAGAAAATGTGACGGTACCGGTACGCA 1021
 DB 721 TACCACTACAGGCAATTCCTTGGCTTAAGAGAAAATGTGACGGTACCGGTACGCA 780
 QY 1022 TTTGTTCAATGGATCCCTCCGTCATCATAGGCTCCCGGACGGTGGGCGCTTGTCT 1081
 DB 781 TTTGTTCAATGGATCCCTCCGTCATCATAGGATCCCGGACGGTGGGCGCTTGTCT 840
 QY 1082 TGTGCGGACGTCGCGGGACGCAATGTTGAGTTGATCGCGGCGGTGCAATTTGGC 1141
 DB 841 TGTGCGGACGTCGCGGGACGCAATGTTGAGTTGATCGCGGCGGTGCAATTTGGC 900
 QY 1142 CGATGAAAGGTCGATCTAGCTCGTTCCTCAACAGCGCTTTTTCATCAGCGGACCATGA 1201
 DB 1201 CGATGAAAGGTCGATCTAGCTCGTTCCTCAACAGCGCTTTTTCATCAGCGGACCATGA

RESULT 11

RAAF71567
 ID AAF71567 standard; DNA; 1287 BP.
 XX AAF71567;
 XX 30-APR-2001 (first entry)
 XX Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:415.
 DE Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.
 XX
 OS Corynebacterium glutamicum.
 XX WO200100844-A2.
 XX 04-JAN-2001.
 XX 23-JUN-2000; 2000WO-IB00943.
 XX 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99US-0143208.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.

PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX (BADI) BASF AG.
 XX
 XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;
 XX WPI; 2001-061975/07.
 DR P-PSDB: AAB79450.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 3; Page 739-740; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (ii) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins
 CC (iii) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diptheriae in a subject. (i), (ii), (iii) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SQ Sequence 1287 BP; 325 A; 349 C; 326 G; 287 T; 0 other;
 Query Match 20.7%; Score 1237.4; DB 22; Length 1287;
 Best Local Similarity 97.6%; Pred. No. 0;
 Matches 1256; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 242 AACAGCAGGCTCAAGTCCGAGATTAATTAACCTAGATCCGTAGACATAAGACATAC 301
 DB 1 AACAGCAGGCTCAAGTCCGAGATTAATTAACCTAAATCCGTAGACATAAGACATAC 60
 QY 302 GTCCTATGCTTCGTGGAAGAACCAATAACCTCAGAAAGATGCGAAGTGGTGCAATTA 361
 DB 61 GTCCTATGCTTCGTGGAAGAACCAATAACCTCAGAAAGATGCGAAGTGGTGCAATTA 120
 QY 362 TCAGAAAATCCAGGTCAAGAGTTAAAAAATGAGGGAAGAAATTTCCCCCTCGG 421
 DB 121 TCAGAAAATCCAGGTCAAGAGTTAAAAAATGAGGGAAGAAATTTCCCCCTCGG 180
 QY 422 GGTGATTGATGGCTTTCTCAACCTCGAAACGGGATCATCACGGAATCTCTGGAGAAC 481
 DB 181 GGTGATTGATGGCTTTCTCAACCTCGAAACGGGATCATCACGGAATCTCTGGAGAAC 240
 QY 482 AGCACCTTAAACCGAGATTCCACCCGAACTCCCAACGATGTTCCCGGTTTATTGA 541
 DB 241 AGCACCTTAAACCGAGATTCCACCCGAACTCCCAACGATGTTCCCGGTTTATTGA 300
 QY 542 TCTTCATAATCACCGTGGAAACGGTGGCGGTTTCTTACGGGAACGAGCAGGCGGAG 601
 DB 301 TCTTCATAATCACCGTGGAAACGGTGGCGGTTTCTTACGGGAACGAGCAGGCGGAG 360

QY 602 GAACACCGCGCAGTATACCGCGAACAATGGCAGCACCGTGAATGTTGCCAAGCATGGTTTC 661
 DB 361 GAATCGCGCGCAGTATACCGCGAACAATGGCAGCACCGTGAATGTTGCCAAGCATGGTTTC 420
 QY 662 GCGCGCGCTGACGACCTGGCAGCGCAGTGGAACCTTATTCCTTGTGTGAAGAGT 721
 DB 421 GCGCGCGCTGACGACCTGGCAGCGCAGTGGAACCTTATTCCTTGTGTGAAGAGG 480
 QY 722 CCTGCTGTGCGGCATTACCTCGAGGGCCCTTTCATCAACGCATCCGCTGTGTGTCCTCA 781
 DB 481 CCTGCTGTGCGGCATTACCTCGAGGGTCTTTCATCAACGCATCCGCTGTGTGTCCTCA 540
 QY 782 AAACCCGGATTTCATTTTCCCGGCAACCCACACAGATCTTCCCGGGTGATCCATGCGGG 841
 DB 541 AAACCCGGATTTCATTTTCCCGGCAACCCACACAGATCTTCCCGGGTGATCCATGCGGG 600
 QY 842 AAAAGTTTGGATCAATTCGATCAGCTAGCGCGGAAATGACAAATCTTTTGTGACTTCT 901
 DB 601 AAAAGTTTGGATCAATTCGATCAGCTAGCGCGGAAATGACAAATCTTTTGTGACTTCT 660
 QY 902 CGATCTCTCGCAGCGCACCATCATGCTTCCCTTCGGGCACATGATGAGATTTGA 961
 DB 661 CGATCTCTCGCAGCGCACCATCATGCTTCCCTTCGGGCACATGATGAGATTTGA 720
 QY 962 TACCCTACACGCGCAATTCCTTGGCTAAAGAGAAAATGTGACGGTCACGGCTACGCA 1021
 DB 721 TACCCTACACGCGCAATTCCTTGGCTAAAGAGAAAATGTGACGGTCACGGCTACGCA 780
 QY 1022 TTTGTTCAATGGATGCTTCGCGTCGATCATAGGCTCCCGCAGCGTGGGCGCTTGGT 1081
 DB 781 TTTGTTCAATGGATGCTTCGCGTCGATCATAGGATCCCGCAGCGTGGGCGCTTGGT 840
 QY 1082 TGTGCGGCACGTGCGGGGACGCATATGTTGAGTTGATGCGCGACGGCGTGCATTTGGC 1141
 DB 841 TGTGCGGCACGTGCGGGGACGCATATGTTGAGTTGATGCGCGACGGCGTGCATTTGGC 900
 QY 1142 CGATGAAACGGTGCATGCTAGCTGTTTCCAAACGCCCTTTTTCATCAGGAGCGCATGGA 1201
 DB 901 CGATGAAACGGTGCATGCTAGCTGTTTCCAAACGCCCTTTTTCATCAGGAGCGCATGGA 960
 QY 1202 AGCGCGCGGAATGCCAGACGGTGAGTACATTTTGGCGGTTTGAACGTCAACGTCAACCGA 1261
 DB 961 AGCGCGCGGAATGCCAGACGGTGAGTACATTTTGGCGGTTTGAACGTCAACGTCAACCGA 1020
 QY 1262 TGGAGTGCCTGCTGCGCGATGGCGGCGCCATCGCGGGGGCACCACACACTAGCGAG 1321
 DB 1021 TGGCGTGCCTGCTGCGCGATGGCGGCGCCATCGCGGGGGTACCCAGCAGTACGCGAG 1080
 QY 1322 TCAGTTTCGTGCACCGTCCGAGGGGTATGAGCTTATCGACGGGACCCCTCCACACCTC 1381
 DB 1081 TCAGTTTCGTGCACCGTCCGAGGGGTATGAGCTTATCGACGGGACCCCTCCACACCTC 1140
 QY 1382 AACCGTCGCGCTAAATTTCTCGTCTTGGCGATCAGAAATCGGTAATCAACACCCCTGC 1441
 DB 1141 AACCGTCGCGCTAAATTTCTCGGACTTAGCGATCAGAAATCGTAAATCAACACCCCTG 1200
 QY 1442 AAATTTGTTGGTCTTGTACTCAACGGCCAGGTGGAAGGTCCTATTAGGTCATCAAGT 1501
 DB 1201 AAATTTGTTGGTCTTGTACTCAACGGCCAGGTGGAAGGTCCTATTAGGTCATCAAGT 1260
 QY 1502 ACTTTAAGTACGAGTAAATCTATCTCTG 1528
 DB 1261 AATTTAAATACGAGCAAACTTTCTCTG 1287

RESULT 12

AAH67872

ID AAH67872 standard; DNA; 1152 BP.

XX AC AAH67872;

XX DT 26-SEP-2001 (first entry)


```

PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
XX P-PSDB; AAB79454.
XX
DR New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 3; Page 751-752; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (ii) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (i) or SMP proteins
CC (iii) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diptheriae in a subject. (i), (ii), (iii) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 882 BP; 259 A; 232 C; 213 G; 178 T; 0 other;
Query Match 12.8%; Score 762; DB 22; Length 882;
Best Local Similarity 93.0%; Pred. No. 3.6e-219;
Matches 820; Conservative 0; Mismatches 50; Indels 2; Gaps 2;
QY 1449 GTGCTCTTTGACTCAAGCGCCAGTGCAGAAAGTCCATTAGTCTCATCAAGTACTTTAA 1508
DB 1 GTGCTCTTTGACTCAAGCGCCAGTGCAGAAAGTCCATTAGTCTCATCAAGTACTTTAA 60
QY 1509 GTACAGTAAACTATCTGTAT-TTTAAAGAGTCCCAACCATGGAATCACTATCTGCAA 1567

```

```

Db 61 ATACGAGCAAAACTTTCCTGATAATAAAGGAGTCCGACCATGGACATCATCATCTGCAA 120
QY 1568 AGACGAGCAAGAAGTCGCAAAAGCAGTTGCGAGTTCCTAATCGCACCTTTGCGCAACAAGG 1627
Db 121 AGACGAGCAAGAAGTCGCAAAAGCAGCGCCCTGATCGCACCTTTGCGCAACAAGG 180
QY 1628 TGGAACTTTGGGGCTTGCNACAGGATCTCACCATTGAGTACCTACCAAGAGCTCATTCG 1687
Db 181 CGGAACCTTTGGGGCTTGCNACAGGATCTCACCATTGAGTACCTACCAAGAGCTCATTCG 240
QY 1688 CATGTATGAAGCTGGGGAAGTGTCAAGAACTGCAAGGCAATCTCTTTGGTGAATA 1747
Db 241 CATGTATGAAGCTGGGGAAGTGTCAAGAACTGCAAGGCAATCTCTTTGGTGAATA 300
QY 1748 COTGGACTAACCCCGTGGAGGATGAAGAGCTACTTTAAACCATTCGCAAGAGTTTCA 1807
Db 301 COTGGATTAACCGCGCAGGATGAAGAGCTACTTTAAACCATTCGCAAGAGTTTCA 360
QY 1808 TGACCACATCGACATCGTTGATGAAGAGTCTACAGCCAGATGTTGCAAAACCTTGATCC 1867
Db 361 TGACCACATCGACATCGTTGATGAAGAGTCTACAGCCAGATGTTGCAAAACCTTGATCC 420
QY 1868 ATACGAAGCAGCTGCAGAGTATGAGCAAGATGCTGCAGAAATCCGTTGAAGTTCAAT 1927
Db 421 ATACGAAGCAGCTGCAGAGTATGAGCAAGATGCTGCAGAAATCCGTTGAAGTTCAAT 480
QY 1928 COTTGATCGCGGGAAGCGG-CAGATCGCTTTTCAATGAAGCATCATCTCTCTGTCAGG 1986
Db 481 COTTGATCGCGGGAAGCGGCAATCGCTTTCAATGAAGCATCATCTCTCTGTCAGG 540
QY 1987 ACTGACAAAGTCCAGGGCTGCACCTTAAACTGTGGAGGACAAACGCTCGATTTCAA 2046
Db 541 ACTGACAAAGTCCAGGGCTGCACCTTAAACTGTGGAGGACAAACGCTCGATTTCAA 600
QY 2047 CACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGCACTTTGTCGCGGC -2106
Db 601 CACCATCGAAGAGTCCCAACCCAGCGCTCACCCAGGTTTGGCACTTTGTCGCGGC 660
QY 2107 GCAAAACATCGTGTGTGGTGGCAACTGGTGAAGGAAAGCCGACGCAATCCGCGGAACGT 2166
Db 661 GCAAAACATCGTGTGTGGTGGCAACTGGTGAAGGAAAGCCGACGCAATCCGCGGAACGT 720
QY 2167 GAAGGCCACAGTACTGCTTCTTCCAGGTTCCATCTGTAGATGCACACATGCCACC 2226
Db 721 GAAGGCCACAGTACTGCTTCTTCCAGGTTCCATCTGTAGATGCACACATGCCACC 780
QY 2227 ATCATCGTTGGATGAAGCAGCAGTATCCAAAGCTGGAACCGCTGATCACTACCGTCTCAT 2286
Db 781 CATCATCGTTGATGAAGCAGCAGTATCCAAAGCTGGAACCGCTGATCACTACCGTCTCAT 840
QY 2287 GGAGCAATTAAGTCCGCTAGAAACAAAGGAAAGTACTG 2328
Db 841 GGAGCAATTAAGTCCGCTAGAAACAAAGGAAAGTACTG 882
RESULT 15
AAH67871
ID AAH67871 standard; DNA; 759 BP.
XX
AC AAH67871;
XX
XX 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2906.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN BP1108790-A2.
XX

```

PD 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127588.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR P-PSDB: AAG2652.
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 8; SEQ ID NO: 2906; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 759 BP: 211 A; 209 C; 189 G; 150 T; 0 other;
SQ
Query Match 11.3%; Score 671.8; DB 22; Length 759;
Best Local Similarity 93.7%; Pred. No. 6.4e-192;
Matches 711; Conservative 0; Mismatches 47; Indels 1; Gaps 1;
1548 ATGGAATCACTCTCTGCAAGACGAGCAGCAAGAGTGGCAAGCACTTGCAGTCTTAATC 1607
D 1 ATGGACATCATCTCTGCAAGACGAGCAGCAAGAGTGGCAAGCACTTGCAGTCTTAATC 60
1608 GCACCTTCGCCAACAGGCTGGAACCTTGGGCTTGCACAGGATCCTCACCACTGAGT 1667
D 61 GCACCTTCGCCAACAGGCTGGAACCTTGGGCTTGCACAGGATCCTCACCACTGAGT 120
1668 ACCTACCAAGAGCTCATTCCGATGTATGAAGCTGGGAGTGTCAATTCAGAACTGCAAG 1727
D 121 ACCTACCAAGAGCTCATTCCGATGTATGAAGCTGGGAGTGTCAATTCAGAACTGCAAG 180
1728 GCATCTTGTGGATGAATACGTTGGGCTTGCACAGGATCCTCACCACTGAGTCTTAAT 1787
D 181 GCATCTTGTGGATGAATACGTTGGGCTTGCACAGGATCCTCACCACTGAGTCTTAAT 240
1788 ACCATTCGCAAGAGTCTCACTGACCAATGACATCTTGTGAAGAGGTCTTACAGCCCA 1847
D 241 ACCATTCGTAAGAGTCTCACTGACCAATGACATCTTGTGAAGAGGTCTTACAGCCCA 300
1848 GATGGTCAACACCTGTATCCATACGAGAGCTGCAGAGTATGAGCAAGATCGCTGCA 1907
D 301 GATGGTCAACACCTGTATCCATACGAGAGCTGCAGAGTATGAGCAAGATCGCTGCA 360
1908 GAATCCCGTTGAATCTCAATCCCTTGGCATCGGCGGAACGG-CACATCGCTTTTCATTGAA 1966
D 361 GAATCCCGTTGAATCTCAATCCCTTGGCATCGGCGGAACGGCCACATCGCTTTCAATGAG 420
1967 CCATCATCTTCTCTGTGAGGAGTCAAAAGGTCCAGGCGCTGACCCCTAAACTGTGGAG 2026
D

Db 421 CCATCATCTTCTCTGTCTAGGACTGACAAAGGTCAGGGCTGCACCTTAATACTGTGGAG 480
QY 2027 GACAAAGCTCGATTCTTCAACACCATCGAAGAGGTCCCAACCCACGCGTCAACCCAGGT 2086
Db 481 GACAAAGCTCGATTCTTCAACACCATCGAAGAGGTCCCAACCCACGCGTCAACCCAGGT 540
QY 2087 TTGGGCACTTTTCCCGCGCGCAAAACATCGTGTGGTGGCAACTGGTGAAGGAAAGCC 2146
Db 541 TTGGGCACTTTTCCCGCGCGCAAAACATCGTGTGGTGGCAACTGGTGAAGGAAAGCC 600
QY 2147 GACGCCATCCGCGGAACCTGTGAAGGCCACCTGCTTCTTCCAGGTTCATCTG 2206
Db 601 GACGCCATCCGCGGAACCTGTGAAGGCCACCTGCTTCTTCCAGGTTCATCTG 660
QY 2207 TAGATGCAACATCCGCGGAACCTGTGAAGGCCACCTGCTTCTTCCAGGTTCATCTG 2266
Db 661 CAGATGCAACATCCGCGGAACCTGTGAAGGCCACCTGCTTCTTCCAGGTTCATCTG 720
QY 2267 GCTGATCACTACCGTCTCATGAGCAATTAAGGTGCGC 2305
Db 721 GCTGATCACTACCGTCTCATGAGCAATTAAGGTGCGC 759

Search completed: September 30, 2003, 19:21:41
Job time : 939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 18:38:12 ; Search time 7503 Seconds

(without alignments)
19335.388 Million cell updates/sec

Title: US-10-019-284A-1

Perfect score: 5969

Sequence: 1 agcctcgagccaccatt.....tgatatatttgcgcgcgtgaa 5969

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pla:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	174.8	2.9	917	29	BZ685850
C 2	123.6	2.1	890	14	CD499281
3	123.6	2.1	1030	14	CD499280
4	103.8	1.7	818	12	BM017859

5	103.2	1.7	669	9	AL656223
6	102.6	1.7	881	13	EX458246
7	102.6	1.7	947	13	EX374555
8	102.6	1.7	1051	12	BM476096
9	102.6	1.7	1063	12	BM910583
10	102.6	1.7	1095	9	AL560879
11	102.6	1.7	1201	13	EX406188
12	102.6	1.7	1201	13	EX421863
13	102.4	1.7	1040	29	CNSOINFL
14	102.2	1.7	669	10	BE790435
15	102.2	1.7	847	13	BU171522
16	101.8	1.7	1201	9	AL546345
17	101.4	1.7	1201	13	EX324827
18	101.2	1.7	1201	13	EX428009
19	100.2	1.7	875	13	BU158666
20	100	1.7	746	12	BI758774
21	100	1.7	812	12	BI825295
22	100	1.7	905	13	BU179037
23	100	1.7	934	13	BO929688
24	100	1.7	984	13	BU167870
25	99.6	1.7	712	10	EG431403
26	99.6	1.7	767	10	EG388237
27	99.6	1.7	940	12	BM472534
28	99	1.7	851	14	CB558813
29	99	1.7	901	13	BU904006
30	98.6	1.7	1201	9	AL549852
31	98.2	1.6	776	12	BM016765
32	98.2	1.6	801	13	BU167497
33	98.2	1.6	870	10	BE742200
34	98.2	1.6	1042	12	BM557913
35	97.8	1.6	665	12	BI553285
36	97.8	1.6	757	12	BG771300
37	97.2	1.6	679	12	BG971729
38	96.6	1.6	848	12	BI601355
39	96.6	1.6	942	12	BI760659
40	96.6	1.6	1201	13	EX445414
41	96.4	1.6	775	28	BR398566
42	96.2	1.6	632	9	AL640829
43	95.6	1.6	627	10	BE280307
44	95.6	1.6	647	12	BJ332387
45	95.2	1.6	640	12	BM487491

ALIGNMENTS

RESULT 1
BZ685850/c
LOCUS PUB666TD 2M_0.6_1.0.KB zea mays genomic clone ZM66Ta035X11,
DEFINITION genomic survey sequence.
ACCESSION BZ685850
VERSION BZ685850.1 GI:28245462
KEYWORDS GSS.
SOURCE zea mays
ORGANISM zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (Bases 1 to 917)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick
A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelawetigr.org
Seq primer: TF
Class: Sheared ends.
Location/Qualifiers

```
source 1. .917
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZM0.6.1.0_KB"
/notes="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"
BASE COUNT 197 a 287 c 279 g 154 t
ORIGIN

Query Match 2.9%; Score 174.8; DB 29; Length 917;
Best Local Similarity 52.7%; Pred. No. 3e-39;
Matches 483; Conservative 0; Mismatches 412; Indels 21; Gaps 4;

QY 3930 TGAAGGACCTTTGAACTGGGGCATGTTCCAGATCATGTCGTCGGGCGAGCGATGTGG 3989
Db 914 TCAAGGTTGCTTCTTACCGCGGGCTGTTCAGGTGTGTAATCGCCCGGTGAAGTGG 855
QY 3990 ATCATGTTTCAAGAACTCGATGACGACCACTCCAAAGACATCGTGTGTCCACAGAC 4049
Db 854 AGAAGTCTACACGCCCTCGGGGACGACACCGC---CTCGCGCGCGCCACCATCGCGG 798
QY 4050 AGCTAAAGATGTTGGTAAACACGACCAACTGTTTCAGCGGTGCTGTGAAGGATTTGG 4109
Db 797 ACGTGAAGAAAGGCGCGACAGACCAACGCCCATGAGCGTCTGTGGGGTGTCT 738
QY 4110 CGGACATTTCTGCGCGCTGATTCGAATCTTGGTTGGTGGCGGTCTGCTCATGGCTATCA 4169
Db 737 CGGAGTGTTCATCGCGATCTCCCGCGCTGATCATTCGCGCGCTGTGTGATGGCGGTGA 678
QY 4170 ACAATGTTGTTGGTGGCGAGATCTGT---TCGGTCCGCAATCACATGTTGGAGATTTCC 4226
Db 677 ACAACCTGATGGCGCAAGGCGATGTTTCATGAGGCGCAGAGCGTCTCGAGGCGCATC 618
QY 4227 CTGAGTACGCGTGTGTGAGATGATCAACCTGATGCACTGTCGCGCGTTCGGGTTCT 4286
Db 617 CGAACCTGATGGCTGTGAGGCGCTGATCAACCTGATGGCCACACCTCTGCTGTGTTCC 558
QY 4287 TGGCAGTGTGTTGTTTTCACCGCAACCAAGGTTTCGTTGTCGAATGATCTCTTGGGCG 4346
Db 557 TGGCAGGCGTGTGGCTGTGTCGGCGCAAGGCGTTTGGCGCAGTGAATCTTGGCA 498
QY 4347 CCGCATGTGATGGGATGGTGTTCACCAACCTGTGTTACGCGTACGACGTGGCGCGCA 4406
Db 497 TCGTGTGTCCTGATGCTGTGATCCAGACCTGCTCAAGCCTGGAACCTACGCGCAAG 438
QY 4407 CCATGACCGCGGCGAAATGCCAATGTGTCCTGTTT-----GTTGGATG 4454
Db 437 CGGTGCGGCTCTCGAGCGCCAGAGCTGCGTACTTCGATATTTTCGTTGGTTCAAGA 378
QY 4455 TTGCTCAAGCTGTTACAGGCGACCGTGTCTCTGCTGTGGTGTCTCTTGGATTTCTGG 4514
Db 377 TTGAGAAAGTGGCTACAGGGCGAGATCTGCGGATCTGATGTCGCGCTATGTGATGA 318
QY 4515 CAACGATCGAAGTGTCTCTCAACAGGCTATGCGGCTGCGAGACTTCTTGTATCACCC 4574
Db 317 GCGTATCGAAGATGCTGCGGCGGAGTGCCCAACGGATCAATTGCTGGTCTGTCG 258
QY 4575 CAGTGTGATCTGTGCTCACCGGTTCCTTACGTTCAATGCTATGTTGTCAGCAATGC 4634
Db 257 CGATCACCACTCGTGTACCGGGGTGTCGCGCTGCGCATATTCGCGCGGTGACCC 198
QY 4635 GCTGGTGGTGTACTGTCGACACAGGCTCTGAGGAGCTCTATGATTTGCGGTGTCAG 4694
Db 197 GTCACTTGGCATCTGATCACCAGAGGCGTGTGCTACCTTGTTCGACCTGGCACCCATGG 138
QY 4695 TCGCGGCTGTGCTTGGTCTGTGCTACTACCAATCGTATCACTGCTGTGACCACT 4754
Db 137 TCGCGCGCGGATTTTCGGTCTGCTGTATGCGCGTGTGTGATCAGCGCATGCCACCA 78
QY 4755 CCTTCCCGCCAAATGAGCTGAGC---TGTTCACACAGGTTGATCTTCACTCTCGCA 4811
```

```
Db 77 TGTTCCTGCGGTGAGCTGACGCTTATTTCTACTCAGGGCGCACCTTCTGTCGCGCA 18
QY 4812 CCGCATCCATGGCAA 4827
Db 17 TGATCGTTATGTCCAA 2

RESULT 2
LOCUS CD499281 890 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA38-D12.Y13-s SHGC-CDA Gasterosteus aculeatus cDNA clone
ACCESSION CDA38-D12.3', mRNA sequence.
VERSION CD499281.1 GI:31426312
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
REFERENCE 1 (bases 1 to 890)
AUTHORS Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J. and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
JOURNAL Unpublished
COMMENT Contact: Kingsley, DM
HMNI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingsley@cgm.stanford.edu
Plate: 38
High quality sequence start: 13
High quality sequence stop: 799.
Location/Qualifiers
1..890
/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA38-D12"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/notes="Vector: lambda ZAP Express/pBK-CMV; Site_1: EcoRI
(5' adaptor); Site_2: XhoI (3' linker primer); The mixed
organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dT sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."
BASE COUNT 194 a 229 c 209 g 257 t 1 others
ORIGIN

Query Match 2.1%; Score 123.6; DB 14; Length 890;
Best Local Similarity 52.3%; Pred. No. 3.7e-24;
Matches 346; Conservative 0; Mismatches 309; Indels 7; Gaps 3;

QY 1606 TCGCACCTTTCGCCAACAGAGGTGACCTTGGGCTTCACAGGATCTTCACACTGA 1665
Db 771 TCCGACCTTGGTCCGACAGATTTTACCTTGGGCTTCCACAGGACGCCCAATGG 712
```



```

Db      708  CCCTTTCCAGCAGCAGCCGACACAGATTTTGTGTGATGAAGACGCCACATGGAAAC 767
QY      2259  TG 2260
Db      768  TG 769

RESULT 4
LOCUS   BM017859
DEFINITION 603645335F1 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5427056 5',
mRNA sequence.
ACCESSION BM017859
VERSION   1
KEYWORDS  BM017859.1 GI:16532213
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 818)
JOURNAL   NIH-MGC http://mgc.ncl.nih.gov/.
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Ling Hong/Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          plate: LCM1892 row: e column: 09
          High quality sequence stop: 803.
FEATURES             Location/Qualifiers
     source           1..818
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5427056"
                     /issue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site1: XhoI; Site2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5',
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC library."
BASE COUNT  207 a 206 c 221 g 184 t
ORIGIN
Query Match 1.7%; Score 103.8; DB 12; Length 818;
Best Local Similarity 49.5%; Pred. No. 2.5e-18;
Matches 327; Conservative 0; Mismatches 327; Indels 7; Gaps 2;

QY      1610  ACCCTTCGCCAACAGGGTGGAACTTTGGGTTGGCAACAGGATCCTACCACCTAGTAC 1669
Db      78  ACCCAGGCCAGAGAAGTACTTACCTCGGGGCTCCCTGAGTACCCCACTGGGTG 137
QY      1670  CTACCAAGAGCTATTTCGGCATGTATGAAGCTGGGGAAGTGTATTCAAGAACTGCAAGGC 1729
Db      138  CTACAGAAGCTGATTGAATACTATAAGATGGGACCTGTCTTTAATATGTGAAGAC 197
QY      1730  ATTCTTTGGTGAATACGTGGGACTAACCCGTGACGATGAAGAAACAGACTACTTTAAAC 1789
Db      198  CTTCAACATGGATGAGTAGCTGGGCTTCTCTCGAGACCACCGGAGATTACCACTCCTT 257
QY      1790  CATTCCCAAGAGTTCACTGACCATCGACATCGTTGATGAAGAGGTCTAGACCCAGA 1849

```

```

Db      258  CATGTGGAACAACATTCTTAAAGACACATTGACATCCACCAGAAACACCCACATTTCTGGA 317
QY      1850  TGGTGCRAAACCTTGATCCATACGAGAGAGCTGCAGATATGAGCAAGATCGGTGCAGA 1909
Db      318  TGGGAATGCAGTCGACCTACAGGAGAAATGTGATGCCCTTTGAAGAGAAATCAAGGCTGC 377
QY      1910  ATCCGTTGAAGTTCAAATCCCTTGGCATCGGGG---AAAGGGCACATCGCTTTTCAATGA 1965
Db      378  AGTGGGATCGAGCTATTGTTTGGAGGCATCGGCCCTGATGACACATTTGCTTCAACGA 437
QY      1966  ACCATCATCTTCTGTGTCAGGACTGACAAAGGTCCAGCGCTGCACCCCTAAATCTGGA 2025
Db      438  GCCAGGCTCCAGTCTGTGTCCAGAGCCCGTGTGAAGACGCTGGCCATGGATACCATCT 497
QY      2026  GGACAACGCTCGATTCTTCAA---CACCATCAAGAGAGTCCCAACCCACGCGCTCACCCA 2082
Db      498  GGCCAATGCTAGGTCTTTCGATGGAGAACTCACCAAGGTGCCACCATGGCTTGACGGT 557
QY      2083  GGGTTTGGGCACTTGTCCCGCGGCAAAACATCGTTGTTGGTGGCACTGGTGAAGAAA 2142
Db      558  GGGGTGGGCACCTGTCTATGGATGCTAGAGAGGTGATGCTCTTATCACAGGTGCTCAAA 617
QY      2143  ACCGACGCCCATCCGCGGAAGTGGAGAGGCCCACTGCTCTTCTGCCAGGTTCAT 2202
Db      618  GGCATTGCTCTGTACAAAGGCCATCGAGGAGGAGTGAACACATGTGGACCGTGTCTGC 677
QY      2203  CCTGTAGATGCACAACATGCCACCATCATCTCTGTGATGAAGACGACGATGCCAAGCTGGA 2262
Db      678  CTTCCAGCAGCATCCCGACCGGGTTTGTGTGTGACGAGGATGCCACCTTGGAGCTGAA 737
QY      2263  A 2263
Db      738  A 738

RESULT 5
LOCUS   AL656223
DEFINITION 569 bp mRNA linear EST 13-DEC-2001
ACCESSION AL656223
KEYWORDS  AL656223.1 GI:17668464
SOURCE    Silurana tropicalis (western clawed frog)
ORGANISM  Silurana tropicalis
REFERENCE Huckle.E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
AUTHORS  Sanger Xenopus tropicalis EST project 2001 (10_2001)
TITLE     Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL   Unpublished
COMMENT   Contact: Huckle E
          Sanger Centre
          Hinxton, Cambridgeshire, CB10 1SA, UK
          Email: tropesanger.ac.uk
          Sanger Xenopus tropicalis EST project 2001
          TROPICALIS_SEQUENCE_ID: TNeu035007.SP6
          Sequencing primer: SP6
          This sequence is from a Xenopus Gene Collection (XGC) library
          constructed by Aaron M. Zorn.
          Location/Qualifiers
             1..669
             /organism="Silurana tropicalis"
             /mol_type="mRNA"
             /db_xref="taxon:8364"
             /clone="TNeu035007"
             /dev_stage="neurula"
             /lab_host="Escherichia coli DH10B"
             /clone_lib="XGC-neurula"
             /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
             was oligo dT primed from 5ug of poly A+ RNA from neurula.
             EcoRI-NotI cut cDNA was then ligated into pCS107 with

```


cgi-bin/cluster.cgi?seq=CS0DA012DB10QPl&cluster=3200.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DA012DB10QPl.

FEATURES
Location/Qualifiers
1..881
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA012YD20"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT 224 a 217 c 243 g 191 t 6 others
ORIGIN

Query Match 1.7%; Score 102.6; DB 13; Length 881;
Best Local Similarity 49.8%; Pred. No. 5.8e-18;
Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

1632 ACCCTGGGGCTGGCAACAGGATCCCTCACACTGACTACCTACCAAGAGCTCATTCGCATG 1691
175 ACCCTGGGGCTCCCACTGGGAGTACCCCACTTGGCTGTACAGAAGCTGATTGAATAC 234
1692 TATGAAGCTGGGGAAGTGTCTATCAAGAACTGCAAGGATCTCTTGGTGGATGATACGTG 1751
235 TATAAGAATGGGAGCTGTCTCTTTAATATGTGAAGAGCTTCAACATCGATGAGTACGTG 294
1752 GGACCTAACCCGTGAGGATGAAGAGAGCTTAAACCACTTCGAAAGAGTTCACCTGAC 1811
295 GGCTTCTTCGAGAGACCCCGGAGAGTTACCACTCCTTCATGTGGAACAACTCTTTCAG 354
1812 CACATCGATCGTTGATGAAGAGGTCTACAGCCAGATGTGCAAAACCCTGATCCATAC 1871
355 CACATTCATCCACCCAGAAAACACCCACATCTTCGATGGATGGAATGCAGTGCACCTAC 414
1872 GAGCAGCTGCAGATGATGAGCAAGATCGCTGCGAGAACTCGGTGGAAGTTCAATCCTT 1931
415 GCAGAATGTATGCTTCTTGAAGAGAGATCAAGGTGCGAGTGGGATCGAGCTATTGTT 474
1932 GGCATCGCGCGG---AAACGGCAGATCGCTTTCATTGAACCATCATCTTCTCTGTCAGGA 1987
475 GGAGGCATCGGCCCTGATGGACACATTCCTTCACGAGCCAGCGCTCCAGTCTGCTGTC 534
1988 CTGACAAGGTCAGGCGCTGCACCTTAAACTTGGAGGAGCAAGCTCGATTCTTCAA- 2045
535 AGGACCCGTGTGAAGACGCTGGCCATGGATACCATCTCTGGCCAAATGCTAGGTTCTCGAT 594
2047 --CACCATCGAAGAGGTCCCAACCCAGCCGCTCAACAGGTTTGGGCACTTGTTCGCCG 2104
595 GGAGAATCACCAAGTGCACCATGCGCTGAGCGTGGGGTGGGCACTGTGATGAT 654
2105 GCGCAAAACATCGTGTGGTGGCAACTGTGGAAGAAAGCCGAGCCATCGCGGAACT 2164
655 GCTAGAGAGGTGATGCTTATCAGGTGCTCAAGAGGATTTGCTCTGTACAGGCC 714
2165 GTGGAGGCCAGTGAAGTGAAGTGTCTTCTGCGAGGTTCCTCTAGATGACAAACATGCCA 2224
715 ATCGAGAGGAGTGAACACATGTTGGACCGTGTCTGCTTCCAGCAGCATCCCGCACC 774
2225 CCATCATCGTTGGATGAAGCAGCAGTATCCAGCTGGAA 2263
775 GTGTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 813

RESULT 7
BX374555
LOCUS
DEFINITION BX374555 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens

EcoRI at the 5' end and NotI at the 3' end."
189 a 152 c 162 g 166 t

Query Match 1.7%; Score 103.2; DB 9; Length 669;
Best Local Similarity 52.0%; Pred. No. 3.3e-18;
Matches 282; Conservative 0; Mismatches 253; Indels 7; Gaps 2;

1611 CCCTTGGCCCAAGAGGTGAACCTTGGGGCTTGCACAGGATCTTCACCACTGAGTACC 1670
118 CCACCGCTGATAAATACTTACTTTAGGCTCCCAACAGGAGCACCCTAGGATGC 177
1671 TACCAGAGCTCATTCGCAATGATGAAGCTGGGAGTGTCTATTCAGAACTGCAAGCA 1730
178 TACAAAACACTATAGAGTACCAATAAAGCGAGATCTCTCTCAAGTATGTGAAGAA 237
1731 TTCTTGTGATGAATACGTTGGGACTTAACCCGTGACGATGAAGAACTGATTTAAAC 1790
238 TTCAATATGATGATGATGTGGGCTTCCAGGAGCACCTGAGAGCTATCATTCCTTC 297
1791 ATTCCGAAGAGTCACTGACCACTGACATCGTGTGATCAAGAGTCTACAGCCAGAT 1850
298 ATGTGAAACAACTTCTTAAGCACATCGACATCAGCGCTGAGAAATGCCACATCTTGGAT 357
1851 GGTGCAAAACCTGATCCATACGAGGAGCTGCAAGTATGAGGCAAGATCGCTGCAGAA 1910
358 GGAATGCCACAGACCTTCAGGAGAGTGTGACCTGTTTGAAGAAAGATTGGGGTCT 417
1911 TCCGTGGAAGTTCAAATCTTGG---CATCGCGGGAAGAGGACATCGCTTTCATTTAA 1966
418 GGTGGGATGAGTGTGTTGTGAGGATTTGTCCTGATGGCCACATGCTTTTAATGAG 477
1967 CCATCACTTCTCTGTCAGGAGTGAAGAGTCCAGCGGTGCACCTTAAACCTGTTGGAG 2026
478 CAGGTCAGTCTGCTCCAGAGCCCGAGTAAACATTTGCCATGATACCATCTCTG 537
2027 GACAACCTCGATCTTT---CACACCATGCAAGAGGTCCCAACCCAGCCGCTCAACCCAG 2083
538 GCAATGACGGTCTTTGACGCAACCTCTCCAAAGTGCCAAAGATGGCCCTGACTGTA 597
2084 GGTTTGGGACTTTGTCGCCGCGCAAAACATCGTGTGGTGCACACTGGTGAAGGAAA 2143
598 GTGTGGGAGTGTGATGAGTCAAGAGTCAATGATTTCTAATCACTGAGGCCATATA 657
2144 GC 2145
658 GC 659

RESULT 6
BX458246
LOCUS
DEFINITION BX458246 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CS0DA012YD20 5-PRIME, mRNA sequence.

ACCESSION BX458246
VERSION BX458246
KEYWORDS EST, GI:31032960
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 881)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3200.r For
more information about this cluster, see
http://www.genoscope.cns.fr/

cDNA clone CSODB007YEL3 5-PRIME, mRNA sequence.

ACCESSION BX374555
VERSION BX374555.1 GI:30456323

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 3200.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODB007AC07QPlcluster=3200.r. Contact :

Feng Jiang Email: fjiang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODB007AC07QPl.

Location/Qualifiers

1..947

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODB007YEL3"

/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"

/clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 255 a 230 c 233 g 209 t

ORIGIN

Query Match 1.7%; Score 102.6; DB 13; Length 947;

Best Local Similarity 49.8%; Pred. No. 6.1e-18;

Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

YQ 1632 ACCTGGGGCTGGACAGGATCCCTCAGCAGTACCTACCAAGAGCTCATTCGCATG 1691

DB 187 ACCCTGGGCTCCCACTGGGAGTACCCACCTGGCTGCTCAAGAGCTGATGAATAC 246

YQ 1692 TATGAAGCTGGGAAGTGTCAATCAAGAACTGCAAGGATCTTTGTTGGATGAATACGTG 1751

DB 247 TATAAATGGGACCTGTCCCTTTAAATATGTGAAGACCTTCAACATGGATGAGTACGTG 306

YQ 1752 GGACTAACCGTGACGATGAAGAAACAGCTACITTAACACCATCTCGCAAGAGTTCACTGAC 1811

DB 307 GGCCCTCCGAGACCCCGGAGAGTACCACCTCTCATGTGGAAACACTTCTTCAAG 366

YQ 1812 CACATCGACATCGTTGATGAGAGGTCTACAGCCAGATGTGCAACCCCTGATCCATAC 1871

DB 367 CACATGACATCCACCAGAAACACCCACATCTTGATGGAATGCAGTGCACCTACAG 426

YQ 1872 GAAGCAGCTGAGATGATGAGCAAGATGCGTGCAGAAATCGGTTGAAGTTCAATCCTT 1931

DB 427 GCAGATGTGTATGCCTTTTGAAGAGAGATCAAGGCTGCGAGTGGGATGAGTATTTGTT 486

YQ 1932 GCATCGGGGG---AAAGGGACATCGTTTTCATGAACCATCATCTTCTCTGACAGA 1987

DB 487 GGAGGATCGGGCTGTATGAGACATATGCTTCAACGAGCCAGGCTCCAGTCTGGTGTC 546

YQ 1988 CTGCAAAAGGTCCAGCGCTGCACCCCTAAACTGTGGAGGACAGCTCGATCTTCTCAA- 2046

DB 547 AGGACCCGTGTGAAGACGCTGCCATGGATACCATCTCTGGCCATGCTAGTCTTCGAT 606

YQ 2047 --CACATCGAGAGGTCCCAACCCACGCGCTACCCAGGGTTTGGGACATTTGTCGCCG 2104

607 GGAGAACTCACCAGGTGCCACCATGGCCTTGACGGTGGGGTGGGCACCTGTCATGGAT 666

2105 GCGCAAAACATCGTGTGGTGGCAACTGGTGAAGAAAGCCGACGCCATCCCGGAAC 2164

667 GCTAGAGAGTGATGATCTTATCACAGGTGCTCAAGAGCATTTGCTCTGTACAGGCC 726

2165 GTGGAAGCCAGGTGACTGCTTCTTCCAGGTTCATCTGTAGATGCACACATGCCA 2224

727 ATCAGAGGAGGTGACCAACATGTGACCGTGTCTGCTTCCAGGAGCATCCCGCAC 786

2225 CCATCATCGTTGGATGAAGCAGCAGTATCCAACTGAA 2263

787 GTGTTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 825

RESULT 8
BM476096

LOCUS

DEFINITION BM476096 1051 bp mRNA linear EST 05-FEB-2002
AGENCOURT_6480896 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5580152

5' RNA sequence.

ACCESSION BM476096

VERSION BM476096.1 GI:18525138

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12337 row: p column: 09

High quality sequence stop: 648.

FEATURES

Location/Qualifiers

1..1051

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5580152"

/tissue_type="retinoblastoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_67"

/note="Organ: eye; Vector: pCMV-SPORT6; Site1: NotI;

Site2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Library constructed by Life

Technologies."

BASE COUNT 280 a 253 c 289 g 229 t

ORIGIN

Query Match 1.7%; Score 102.6; DB 12; Length 1051;

Best Local Similarity 49.8%; Pred. No. 6.5e-18;

Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

YQ 1632 ACCTGGGGCTGGCAACAGGATCCCTCAGCAGTACCTACCAAGAGCTCATTCGCATG 1691

DB 129 ACCCTGGGCTCCCACTGGGAGTACCCACCTGGCTGCTCAAGAGCTGATGAATAC 188

YQ 1692 TATGAAGCTGGGAAGTGTCAATCAAGAACTGCAAGCATTCTTGTGGATGAATACGTG 1751

DB 189 TATAAAGTGGGACCTGTCTTTAAATATGTGAAGACCTTCAACATGGATGAGTACGTG 248

YQ 1752 GGACTAACCGTGACGATGAAGAAACAGCTACTTTAAACACCATCTGCAAGAGTTCACTGAC 1811

DB 249 GGCCTTCTCGAGACACCCCGGAGAGTTACCACTCTCTTCATGTGGAACAACACTTCTTCAAG 308

QY 1812 CACATCGACATCGTTGATGAAGAGGTCTACAGCCCGAGATGGTGCAAAACCTGATCCATAC 1871
 Db 309 CATTGATGACATCCACCAGAAACACCATCTCTGGATGGGAATCGATCGACCTTACAG 368
 QY 1872 GAAGCAGCTCAGAGTATGAGCAAGATCGCTGCAGATCCGTTGAAGTTCAAATCCTT 1931
 Db 369 GCAGATGTGATGCCCTTGNAGAGAGATCAAGCTGCGAGTGGGATCAGCTATTGTT 428
 QY 1932 GGCATCGGGCGG---AAAGCGCACATCGCTTTCATTGAACCATCATCTTCTCTGTCAGGA 1987
 Db 429 GGAGGATCGGCCCTGATGACACATTCGCTTCAAGGAGCCAGCTCCAGTCTGTGTCC 488
 QY 1988 CTGACAAAGTTCGAGCGCTGCACCTCAAACTGTGGAGACAACCTCGATTCTTCAA- 2046
 Db 489 AGACCCGCTGAGAGCGCTGGCCATGGATACCATCTCTGCCAATGCTAGGTTCTTCGAT 548
 QY 2047 --CACCATGAAGAGTCCCAACCCAGCGCTGACCCAGGGTTGGGCACTTTGTCGCCG 2104
 Db 549 GGAGAACTACCAAGAGTGGCCCATGGCTTGACGGTGGGGTGGGCACTGTCATGGAT 608
 QY 2105 GCGCAAACTCTGTTGTGGCACTGCTGAGAGGAGAAAGCCGACCCATCCCGGAAT 2164
 Db 609 GCTAGAGAGTGTATGCTTATACAGGTGCTTCAAGGCATTTCTCTGTACAAAGGCC 668
 QY 2165 GTGAAGGCCAGTACTGCTTCTGCCAGGTTCCATCTGATGATGACACATGCCA 2224
 Db 669 ATCGAGGAGGAGTGAACACATGTGGACCGTGTCTGCCCTTCCAGCAGCATCCCGCACC 728
 QY 2225 CCATCATCTGTGATGAGCAGCAGTATCCAGCTGGAA 2263
 Db 729 GTGTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 767

RESULT 9
 BM910583 1063 bp mRNA linear EST 12-MAR-2002
 LOCUS
 DEFINITION AGENCOURT_6703024 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454094
 5', mRNA sequence.

BM910583
 BM910583.1 GI:19360962

EST.
 Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1063)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Cloning by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LCM948 row: k column: 23

High quality sequence stop: 648.

Location/Qualifiers

1..1063

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5454094"

/tissue_type="astrocytoma grade IV, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_98"

/notes="Organ: brain; Vector: pOTB7; Site: 1; XhoI; Site: 2;

EcoRI; cDNA made by oligo-dT priming, Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 280 a 252 c 275 g 249 t 7 others
 ORIGIN

Query Match 1.7%; Score 102.6; DB 12; Length 1063;
 Best Local Similarity 49.8%; Pred. No. 6.5e-18;
 Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;

QY 1632 ACCTTGGGCTGTGCAACAGGATCCTCACACTGACTTACCTACCAAGAGCTCATTCGCGATG 1691

Db 97 ACCTTGGGCTGTGCAACAGGATCCTCACACTGACTTACCTACCAAGAGCTCATTCGCGATG 156

QY 1692 TATGAAGTGGGGAAGTGTCTATCAAGAACTGCAAGCACTTCTTGTGGATGAATACGTG 1751

Db 157 TATAAGATGGGGAAGTGTCTTAAATATGTAAGAACTTCAACATGGATGAGTACGTG 216

QY 1752 GGACTAAACCGTGAGCATGAAACAGCTACITTAACACCATTCGCAAGAGTTCACCTGAC 1811

Db 217 GGCCTTCCTCGAGACCCCGGAGAGTTACCACTCTCTCATGTGGAAACAATCTTCTCAAG 276

QY 1812 CACATCGACATCGTTGATGAAGAGTCTACAGCCCGAGATGGTGCAAAACCTGATCCATAC 1871

Db 277 CACATGATCCACCCAGAAACACCCACATTCGTGATGGATGCAAGTGCAGTGCACCTACAG 336

QY 1872 GAAGCAGCTGAGATGATGAGCAAGATCCTGAGAGTCCGTTGAAGTTCATATCCTT 1931

Db 337 GCAGAATGTGATGCTTGAAGAGAGATCAAGGTGCGAGTGGGATCGAGCTATTGTT 396

QY 1932 GGCATCGGGCGG---AAACGGCACATCGCTTTCATTGAACCATCATCTTCTCTGACGA 1987

Db 397 GGAGGCATCGGGCGGCTGATGGACACATTCCTTCAACGAGCAGCGCTCCAGTCTGCTGCC 456

QY 1988 CTGCAAAAGTCCAGGCGCTGCACCTTAAACTGTGAGGACACGCTCGATCTTCTTCAA- 2046

Db 457 AGGACCGGTGTGAAGACCTGGCCATGGATACCATCTTGGCCATGCTAGGTTCTTCTGAT 516

QY 2047 --CACCATCGAAGAGTCCCAACCCAGCGCTGCACCCAGGGTTGGGCACTTTGTCGCCG 2104

Db 517 GGAGAACTCACCAAGTGGCCACCATGCGCTTGAAGTGGGCTGGGCACTGTCATGGAT 576

QY 2105 GCGCAAAACATCGTTGTGGTGGCACTGTGGAAGAAAGCGGACGCCATCCGCGGAACT 2164

Db 577 GCTAGAGAGTGTATGATCCTTATCACAGGTGCTCAAGGGCATTTGCTGTGTACAGGCC 636

QY 2165 GTGGAAGGCCAGTGCATCTTCTTGGCCAGGTTCCATCTCTGATGATGCAACAATGCCA 2224

Db 637 ATCGAGGAGGAGTGAACACCATGCGAGCGTGTCTGCTTCCAGCAGCATCCCGCACC 696

QY 2225 CCATCATCGTTGGATGAAGCAGCAGTATCCAGCTGGAA 2263

Db 697 GTGTTGTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 735

RESULT 10

AL560879

LOCUS

DEFINITION

AL560879 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED

AL560879 Homo sapiens cDNA clone CS0DL005YD04 5-PRIME, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1095)

AUTHORS

TITLE

JOURNAL

Unpublished

AL560879 1095 bp mRNA linear EST 31-MAY-2003
 AL560879 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
 Homo sapiens cDNA clone CS0DL005YD04 5-PRIME, mRNA sequence.

AL560879

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1095)

AUTHORS

TITLE

JOURNAL

Unpublished


```

Db      379 GCAGAAATGATGCTCTTTGAAGAGAGATCAAGGCTGCAGGTGGGATCGAGCTATTGTT 438
QY      1932 GGCATCGGGG---AAACGGACATCGCTTTCATTGAACCATCATCTTCTCTGTCAGGA 1987
Db      439 GAGGAGATCGGCGCTGATGACACATGCTTCAACGACCGAGCTCCAGTGTGTC 498
QY      1988 CTGACAAAGGTCACGCGCTGCACCCCTAAACTGTGGAGGACAAAGCTCGATCTTCAA- 2046
Db      499 AGGACCGGTGTAAGACGCTGGCCATGGATACCATCTCTGGCCAAATGCTAGGTTCTTCGAT 558
QY      2047 --CACCATCGAAGAGGTGCCCAACCCACGCGCTCACCGAGGTTTGGGCACCTTTGTCGCCG 2104
Db      559 GGAGAACTACCAAGTGTCCCACTGCGCTTCAGCGTGGGGTGGGCACTGTCATGAT 618
QY      2105 GCCCAAAACATCGTGTGGTGGCACTGTGTAAGGAAAGCCGACCCATCCGCGGAAT 2164
Db      619 GCTAGAGAGGTGATGATCTTATACAGGTGTCAAGGCAATTTGCTCTGTACAGGCC 678
QY      2165 GTGGAAGGCCAGTACTGCTTCTGCCAGGTTTCCATCTGTAGATGACACACATGCCA 2224
Db      679 ATCGAGGAGGAGTGAACCATGTGGACCGTGTCTGCTTCCAGCAGGATCCCGCCACC 738
QY      2225 CCATCATCGTGTGATGAAGCAGCAGTATCCAAAGCTGGAA 2263
Db      739 GTCTTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 777

RESULT 12
BX421863
LOCUS      1201 bp      mRNA      linear      EST 13-MAY-2003
DEFINITION      BX421863 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA
                  clone CS0DH005YE10 5-PRIME, mRNA sequence.
ACCESSION      BX421863
VERSION        BX421863.1 GI:30651259
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished
COMMENT        Contact: Genoscope
                  Genoscope - Centre National de Sequencage
                  BP 191 91006 EVRY cedex - France
                  Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
                  Library was constructed by Life Technologies, a division of
                  Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
                  http://fulllength.invitrogen.com/InvitrogenCorporation1600
                  Faraday Avenue Genoscope sequence ID : CS0DH005BC05QP1.
                  Location/Qualifiers
FEATURES       source
                  1..1201
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="CS0DH005YE10"
                  /tissue_type="T CELLS (JURKAT CELL LINE)"
                  /cell_line="JURKAT CELL LINE"
                  /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"
                  /note="vector: pCMVSPORT6; 1st strand cDNA was primed
                  with a NotI-oligo(dT) primer. Five prime end enriched,
                  double-strand cDNA was digested with Not I and cloned into
                  the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                  Library was not normalized."
BASE COUNT      311 a 267 c 316 g 268 t 39 others
ORIGIN
Query Match      1.7%; Score 102.6; DB 13; Length 1201;
Best Local Similarity 49.8%; Pred. No. 7e-18;
Matches 318; Conservative 0; Mismatches 314; Indels 7; Gaps 2;
QY      1632 ACCTTGGGGCTTGCAACAGGATCCTCACCAGTACCTACCAAGAGCTCATTCGGCATG 1691

```

```

Db      229 ACCCTGGGCTCCCGACTGGGAGTACCCCACTTGCCTGCTACAAGAAGCTGATTGAATAC 288
QY      1692 TATGAAGCTGGGAGGTGCTATTCAAGAAGCTCAAGGCAATCTTGTGGATGAATAGGTG 1751
Db      289 TATGAAGTGGGACCTGCTCTTTAAATATGTGAAGACCTTCAACATGGATGAGTACGTG 348
QY      1752 GGACTAAGCCGTGACGATGAAGAAAGACTACTTTAAACCATTTGCAAGAGTTCACATGAC 1811
Db      349 GGCCTTCTCGAGACCCCGGAGAGTTACCACTCTTCATGTGGAACAAGCTTCTTCAAG 408
QY      1812 CACATCGACATCGTTGATGAAGAGGTCTACAGCCAGATGGTGCAAAACCTGATCCATAC 1871
Db      409 CACATGACATCCCGCCAGAAACACCCACATTTGGATGGGAATGAGTGCAGTGCACCTACAG 468
QY      1872 GAAGCAGCTGCAGAGTATGAGGCAAGATCCTGCTGAGAAATCCGTTGAAGTTCAATCCTT 1931
Db      469 GCAGAATGTGATGCTTGTAAAGAAAGATCAAGGCTCAGGTGGGATCGAGCTATTGTT 528
QY      1932 GGCATCGGGG---AAACGGCACATCGCTTTCATTGAACCATCATCTTCTCTGCAGGA 1987
Db      529 GGAGCATCGGCCCTGTGAGACACATTCCTTCAACGAGCCAGGCTCCAGTCTGGTGCC 588
QY      1988 CTGACAAAGTCCAGGCGCTGCACCTAAACTGTGAGGACACAGCTCTGATTTCTTCAA- 2046
Db      589 AGGACCCGTGTGAGACGCTGGCCATGGATACCATCTCTGGCAATGCTAGGTTCTTCGAT 648
QY      2047 --CACCATCGAAGAGTCCCAACCCACGCGCTCACCCAGGTTTGGGCACCTTGTGCCGC 2104
Db      649 GGAGAACTCAACCAAGTGCCACCATGTCCTTCAACGAGCCAGGCTCCAGTCTGATGAT 708
QY      2105 GCGCAAAACATCGTGTGTTGGCAACTGTGTAAGGAAAGCGGAGCCATCCGCGGAAT 2164
Db      709 GCTAGAGAGGTGATGATCTTATCACAGGTGTCAAGAGGATTTGCTCTGTACAAGGCC 768
QY      2165 GTGGAAGGCCAGTACTGCTCTTGTGCCAGGTTCCATCTCTAGATGCACAAATGCCA 2224
Db      769 ATCGAGGAGGAGTGAACCATGTGACCGTGTCTGCTCTCCAGCAGCATCCCGCACC 828
QY      2225 CCATCATCGTGGATGAAGCAGCAGTATCCAAAGCTGGAA 2263
Db      829 GTGTTTGTGTGACGAGGATGCCACCTTGGAGCTGAAA 867

RESULT 13
CNS01NF1      1040 bp      DNA      linear      GSS 14-JUN-2001
LOCUS      Anopheles gambiae GSS T7 end of clone 23002 of NotreDame1 library
DEFINITION      from strain PEST of Anopheles gambiae (African malaria mosquito),
                  genomic survey sequence.
ACCESSION      AL152142
VERSION        AL152142.1 GI:7013061
KEYWORDS       GSS.
SOURCE         Anopheles gambiae (African malaria mosquito)
ORGANISM       Anopheles gambiae
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
                  Anopheles.
REFERENCE      1 (bases 1 to 1040)
                  Genoscope.
AUTHORS        Direct Submission
TITLE          Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
JOURNAL        BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
                  - Web : www.genoscope.cns.fr)
REFERENCE      2 (bases 1 to 1040)
                  Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
AUTHORS        Submitted (16-FEB-2000) BMFI, Institut Pasteur, 25, rue du Dr.
TITLE          Roux, Paris 75015, France
JOURNAL        This clone is from an A. gambiae BAC library provided by F.H.
                  Collins and sequenced by Genoscope in collaboration with the
                  Laboratory of Biochem. and Biol. Molec. of Insects, Institut
                  Pasteur.
COMMENT

```

FEATURES		Location/Qualifiers		Location/Qualifiers	
source		1. .1040		194 a 314 c 322 g 206 t 4 others	
		/organism="Anopheles gambiae"			
		/mol_type="genomic DNA"			
		/strain="PEST"			
		/db_xref="taxon:7165"			
		/clone="23002"			
		/clone_lib="Notredame1"			
		/note="end : T7"			
BASE COUNT		194 a 314 c 322 g 206 t 4 others			
ORIGIN					
Query Match		1.7%; Score 102.4; DB 29; Length 1040;			
Best Local Similarity		49.0%; Pred. No. 7.4e-18;			
Matches 396; Conservative		0; Mismatches 396; Indels 16; Gaps 4;			
QY	3819	TTGGCGGCGAAGACACATGTGCGCGCGACACTGTGCAACGCGTTTACGCCCTCGGC	3878		
Db	59	TTGGCGGCGCGAGAACATCGCCACCGTCAGCCATTGCTATCCCGCGCTTGGTTCGCC	118		
QY	3879	TCAAGACACCAAGGATGTGGATCGCCAAAGTCTGGATGATGATCCAGATCTGAAGGCA	3938		
Db	119	TCAATGACCCAGCAGCGCCAGCCCAAGAAATGGAAGAGCTGCCGATGGTCAAGGTT	178		
QY	3939	CCTTTCAAACTGGCGCATGTCCAGATCATCTGTGGCGCAGCGCATGTGGATCATGTTT	3998		
Db	179	GCTTCACCAACCGCGGCGAGTTCACAGTGTGTGATCGGCACCGACGTCGCGGACTA	233		
QY	3999	TCAAGAACTCGATGACGCAAGCTCCAAAGACATGCTGTCCACAGACAGCTCAAG	4058		
Db	234	CTATCAGCGCTGATCGCCAGCAGCGCGGTCAACGAGCCGACAAAGACAGGCCAAG	292		
QY	4059	ATGTTGTGCTAACAGGCCACTGTTTCAGCGCTGCTGTGAAGTATTGGCGGACATTT	4118		
Db	293	TGCGCGCGCGCAGAACATGACTGTGACGCGCGCACCATTCTCCTCCTCGCGGAGTCT	352		
QY	4119	TGCTCCGCTGATTCACATCTTGGTGTGGCGGTCTGCTCATGCTATCAACAATGTGT	4178		
Db	353	TCTTCCGCTGCTGCCAGCGCTGATCAGCGCGGCTGATCTCGCTTCGCAAGTGA	412		
QY	4179	TGTTTGGGAGGATCTGTTGCTGCTCCCATCTACTGTGAGATGTTTCCCTCAGATCAGG	4238		
Db	413	TCGGCGACATCCCGATGTCGGCGGCGCAACGCTGCGCAGATGCAACCGCGGTGA	472		
QY	4239	GTGTTGCTGAGATGATCAACCTGATGCTGCGCGCTGCTGCTTTCGAGTGTGG	4298		
Db	473	CCATTTACGACTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	532		
QY	4299	TTGGTTTACCGCAACCAAGCGTTTCGGTGGCAATGAGTTCCTGGCGCGCGCATGGTA	4358		
Db	533	TCTGCTGCTGCGAGGTGAAAGATGGCGGCGCACCG-GTGTGCTGCGCATCTGCTGG	591		
QY	4359	TGCGGATGTTTCCCAACCTGTTAACGGCTACGAGCTGCGCGCCACCATGACCGCG	4418		
Db	592	TCACCTGTGTG-TCGCCATGCTCTGATGAATCTTCTACTGCTGCGCCAGCAACCG	550		
QY	4419	CGAAATGCAATGTGGTCCCTGTTGGTTGGATGTTGCTCAAGCTGTTTACCGGGCA	4478		
Db	651	G-----AAGTGTGGAATCTCGGCTGCTGCTGATCCAGAAAGTCGCGCTATCAG	702		
QY	4479	CGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4538		
Db	703	AGTGATCCCGTGCATCTCGCAGGATGCGCGTGGCTGGATCGAAACCCCGGCTGA	762		
QY	4539	AGCGACTCATGGCACTGCACTCTCTGATCACCACCGTGTGATCTGCTGCTCACCG	4598		
Db	763	AGATCGTCCCGCACTATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	822		
QY	4599	GCTTCCTAGTTCAATGCTATTGGTCC	4626		
Db	823	TGTTCTTGGCGCAGCACTGATCGGGCC	850		

RESULT 14

BE790435

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE790435

601477188F1 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:3880173 5',

mrna sequence.

BE790435

BE790435.1 GI:10211633

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/DFP/Gazdar

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LHAM9646 row: o column: 22

High quality sequence stop: 564.

Location/Qualifiers

1. .669

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3880173"

/tissue_type="large cell carcinoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_68"

/note="Organ: lung; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.8 kb. Library constructed by Life Technologies."

165 a 172 c 186 g 145 t 1 others

BASE COUNT

ORIGIN

Query Match

1.7%; Score 102.2; DB 10; Length 669;

Best Local Similarity

50.1%; Pred. No. 6.4e-18;

Matches 311; Conservative

0; Mismatches 303; Indels 7; Gaps 2;

QY

1632

ACCTTGGGCTGTCACAGGATCCTCACCCTGAGTACCTACCAAGAGCTCATTCGCGATG

1691

Db

25

ACCTTGGGCTGTCACAGGATCCTCACCCTGAGTACCTACCAAGAGCTCATTCGCGATG

84

QY

1692

TATGAAGCTGGGAGGTGTCATTCAGAACTGCAAGGATTCCTTTGGTGGATGAATAGTG

1751

Db

85

TATAAGATGGGACCTGCTCTCTTTAAATATGTGAAGACCTTCAACATGGATGAGTAGTG

144

QY

1752

GGACTAACCTGAGGATGAACAGCTACTTTAAACCATTCGCAAGAGTTCACCTGAC

1811

Db

145

GGCTTCTCTCGAGACCCCGCGAGATTACCACCTCTTCATGTGGACAACTTCTCAAG

204

QY

1812

CACATCGACATCGTTGATGAAGAGTCTACAGCCCGAGATGGTGCACACCTGATCCATAC

1871

Db

205

CACATTGACATCCACCCAGAAAACACCCACATCTTGGATGGGAATGCAGTCGACCTACAG

264

QY

1872

GAAGCAGCTGAGAGTATGAGCAAGATCGCTGCGAAGATCCGTTGAATTCATATCCTT

1931

Db

265

GCAGATGTGATGCTTTTGAAGAAGATCAAGCTGCAGTGGGATGCGAGCTATTGTT

324

QY

1932

GGCATCGGGG-AAAACGGCAGATCGCTTTTCAATGAACCATCATCTTCTGTCTCAGGA

1987

Db

325

GGAGCATCGGCCCTGTGTGACACATTCCTTCAACGAGCCAGGCTCCAGTCTGTGTCC

384

QY

1988

CTGACAAAGTCCAGCGCTGCACCTTAAACTGTGGAGACACAGCTCGATTCTTCTTCAA

2046

```

Db      385 AGGACCCGTTGTAAGACGCTGGCCATGGATACCATCTGGCCATCTAGGTTCTTCGAT 444
Qy      2047 --CACCATGCAAGAGTCCCAACGCCACGCGTCACCCAGGTTTGGCAGTTTGTCCCGC 2104
Db      445 GGAGAACTACCAAGGTGGCCACCATGGCTTTGACGGTGGGGTGGCCACTGTCATGGAT 504
Qy      2105 GGCCTAAACATCGTTGTTGGCACTGTGGAAGGAAAGCCGACGCCATCCGCGAAGT 2164
Db      505 GCTAGAGAGGTGATGATCTTATCACAGGTGCTCACAGGCATTGCTGTACAGGCC 564
Qy      2165 GTGGAAGGCCAGTACTGCTTCTTCCAGGTTCCATCCTGTAGATGCACACATGCCA 2224
Db      565 ATCGAGGAGGAGTGAACACATGTGGACCGTGTCTGCTTCCAGCAGCATCCCGCAC 624
Qy      2225 CCATCATCGTTGGATGAAGCA 2245
Db      625 GTGCTGTGTGAAGAAGGA 645

RESULT 15
BUI71522
LOCUS      847 bp mRNA linear EST 04-SEP-2002
DEFINITION AGENCOURT_7928634 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6151420
            S., mRNA sequence.
ACCESSION BUI71522
VERSION   BUI71522.1 GI:22685506
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 847)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DMP
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM3487 Row: k Column: 05
            High quality sequence stop: 536.
FEATURES             Location/Qualifiers
     source            1..847
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6151420"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DHI0B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /notes="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."
BASE COUNT      219 a 210 c 226 g 190 t 2 others
ORIGIN
Query Match      1.7%; Score 102.2; DB 13; Length 847;
Best Local Similarity 49.9%; Pred. No. 7.4e-18;
Matches 311; Conservative 0; Mismatches 305; Indels 7; Gaps 2;

Qy      1632 ACCTGGGGTTCGAACAGAGATCCTACCACTAGTACCTACCAAGAGCTCATTCGCATG 1691
Db      99 ACCCTGGGGTCCCGACCTGGGAGTACCCCACTTGGCTGTCTACAGAGCTGATTGAATAC 158
Qy      1692 TATGAAGCTGGGGAAGTGTGATCAAGAGCTGCAAGCATCTTGTGGATGAATACGTTG 1751
Db      159 TATAAGAATGGGGACCTGTCTTTAAATATGTGAGACCTTCAACATGGATGATACGTG 218

```

```

Qy      1752 GGACTAACCCGTGAGGATGAAGAACAGCTACTTTTAAACACCTTCGCAAGAGTTCACTGAC 1811
Db      219 GGCCTTCTCGAGACCCACCCGGAGAGTTTACCACCTCTTCATGTGGACAACTTTCTTCAAG 278
Qy      1812 CACATCGACATCGTTGATGAAGAGGTCTTACAGCCCGAGATGGTGAAGAAACCCCTGATCCATAC 1871
Db      279 CACATTGACATCCACCCAGAAAACACCCACATTCTGGATGGGAATGCAGTCGACCTACAG 338
Qy      1872 GAAGCAGTGCAGATGATGAGCRAAGATCGCTCAGAAATCCGTTGAAGTTCAATCCTT 1931
Db      339 GCAGAATGTATGCTTTTGAAGAGAAATCAAGGCTGCAGGTGGGATCGAGCTATTGTT 398
Qy      1932 GGCATCGCGGG---AAACGGCAGATCGCTTTTCATTGAACCATCATCTTCTCTGTACAGGA 1987
Db      399 GGAGGCATCGGCCCTGATGGACACATTGCCCTTCAAGGAGCCAGGCTCCAGTCTGGTGTCC 458
Qy      1988 CTGACAAAGGTCAGGCGCTGCACCCCTAAACTGTGGAGGACACAGCTCGATTCTTCAA- 2046
Db      459 AGGACCCGTGTGAAGACGCTGCCATGGATACCATCTCTGGCCAATGTAGTCTCTCGAT 518
Qy      2047 --CACCATCGAAGAGTCCCAACGCCACGCGTCACCCAGGTTTGGGCACTTTGTCCCGC 2104
Db      519 GGAGAACTCACCAAGGTGCCACCATGCGCTTGACGGTGGGGTGGGCACTGTCATGGAT 578
Qy      2105 GCGCAAAACATCGTGTGGTGGCAACTGCTGTGAAGAAAGCCGACGCCATCCGCGGAAGT 2164
Db      579 GCTANAGAGGTGATGATCTTTATCACAGGTGCTCAAGAGGATTTGCTCTGTACAGGCC 638
Qy      2165 GTGGAGGCCCACTGCTGCTTCTTCCAGGTTCCATCTGTAGATGCACAAACATGCCA 2224
Db      639 ATCGAGGAGGAGTGAACACACATGTGGACCGTGTCTGCTTCCAGCAGCATCCCGGCACC 698
Qy      2225 CCATCATCGTTGGATGAAGCAGC 2247
Db      699 NGGTTTGTGTGTGACGAGGATGC 721

```

Search completed: October 1, 2003, 01:21:48
Job time : 7510 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 19:02:07 ; Search time 259 Seconds
(without alignments)
10172.262 Million cell updates/sec

Title: US-10-019-284A-1
Perfect score: 5969
Sequence: 1 agtcgctgagccaccatt.....tgatatatttgcgcgctgaa 5969

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/FCIUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3687	61.8	6911	1	US-08-311-174-4
2	172.2	2.9	3895	4	Sequence 4, Appl1
3	154.2	2.6	9769	4	Sequence 201, Appl
4	139	2.3	732	4	Sequence 30, Appl
5	124.8	2.1	714	4	Sequence 1277, Ap
6	108.6	1.8	3615	1	US-09-107-532A-378
7	108.6	1.8	3615	1	Sequence 17, Appl
8	108.6	1.8	3615	1	US-08-920-812-17
9	108.6	1.8	3615	1	Sequence 17, Appl
10	108.6	1.8	3615	1	US-08-921-177-17
11	108.5	1.8	3615	2	Sequence 17, Appl
12	92.4	1.5	1887	4	US-08-362-577C-17
13	92	1.5	2295	4	Sequence 17, Appl
14	90.6	1.5	834	4	US-08-920-828-17
15	90	1.5	1830121	4	US-08-673-190A-3
16	84.4	1.4	4403765	3	Sequence 3, Appl1
17	84.4	1.4	4403765	3	Sequence 1634, Ap
18	84.4	1.4	4403765	3	Sequence 736, Appl
19	73.8	1.2	1284	4	Sequence 1, Appl1
20	69.4	1.2	1752	4	Sequence 1, Appl1
21	65.6	1.1	357	2	Sequence 1, Appl1
22	65.6	1.1	2726	4	Sequence 1, Appl1
23	65.6	1.1	2726	4	Sequence 1, Appl1
24	65.2	1.1	1034	4	Sequence 1, Appl1
25	64.4	1.1	1314	4	Sequence 1876, Ap
26	62.2	1.0	8394	4	Sequence 3157, Ap
27	61.8	1.0	1164	4	Sequence 6, Appli

c	28	61.8	1.0	1209	4	US-09-252-991A-1624	Sequence 1624, Ap
c	29	61.8	1.0	2127	4	US-09-252-991A-1503	Sequence 1503, Ap
c	30	58.2	1.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	31	58.2	1.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	32	58	1.0	325	4	US-08-961-527-351	Sequence 351, Appl
c	33	55.4	0.9	1437	4	US-09-134-001C-2228	Sequence 2228, Ap
c	34	55	0.9	2435	4	US-09-634-238-131	Sequence 131, App
c	35	52.8	0.9	513	4	US-09-134-001C-297	Sequence 297, App
c	36	52.8	0.9	1971	4	US-09-107-532A-1429	Sequence 1429, Ap
c	37	52.6	0.9	1846	4	US-09-634-238-183	Sequence 183, App
c	38	52.2	0.9	1896	4	US-09-107-532A-248	Sequence 248, App
c	39	51.4	0.9	28882	4	US-08-961-527-140	Sequence 140, App
c	40	50.4	0.8	744	4	US-09-134-001C-2556	Sequence 2556, Ap
c	41	50	0.8	612	4	US-09-252-991A-1623	Sequence 1623, Ap
c	42	50	0.8	936	4	US-09-252-991A-1504	Sequence 1504, Ap
c	43	50	0.8	1296	4	US-09-252-991A-1670	Sequence 1670, Ap
c	44	50	0.8	2061	4	US-09-252-991A-1544	Sequence 1544, Ap
c	45	47.2	0.8	1344	1	US-08-844-010-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-311-174-4
; Sequence 4, Application US/083111174
; Patent No. 5556776
; GENERAL INFORMATION:
; APPLICANT: TSUCHIYA, MAKOTO
; APPLICANT: MIWA, KIYOSHI
; TITLE OF INVENTION: SUCRASE GENE DERIVED FROM CORYNEFORM
; TITLE OF INVENTION: BACTERIA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/311,174
; FILING DATE: 23-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 046836/1992
; FILING DATE: 04-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5556776man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-699-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6911 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-311-174-4

Query Match 61.8%; Score 3687; DB 1; Length 6911;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTCCGTGACGCCACCATGATGTTGGTGGTTCACCGAGCTGCGGAGGCTTCTACATCT 60
DB |||||
1 AGTCCGTGACGCCACCATGATGTTGGTGGTTCACCGAGCTGCGGAGGCTTCTACATCT 60
QY |||||
61 ACCTCCCGTCGGCGTGGAGTGGGTCATTACGGTGGGATCACGGCGTGAAGTTGG 120
DB |||||
61 ACCTCCCGTCGGCGTGGAGTGGGTCATTACGGTGGGATCACGGCGTGAAGTTGG 120
QY |||||
121 GAACCCATGGTGTCTTGGGTGAGGGAACGAGTGGGGTGAAGTTTCAAGTG 180
DB |||||
121 GAACCCATGGTGTCTTGGGTGAGGGAACGAGTGGGGTGAAGTTTCAAGTG 180
QY |||||
181 TGTGAGTTTAAAGTTATGCATCATCAGCTTGGAGGCTGAGTAAATCAGTAGACCTG 240
DB |||||
181 TGTGAGTTTAAAGTTATGCATCATCAGCTTGGAGGCTGAGTAAATCAGTAGACCTG 240
QY |||||
241 CAACAGCAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAAGACATCAT 300
DB |||||
241 CAACAGCAGGCTCAAGTCCGAAGATAATTAACCTAGATCCGTAGACATAAGACATCAT 300
QY |||||
301 CGTCTATGCTGCTGGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTCATT 360
DB |||||
301 CGTCTATGCTGCTGGAAGGAACCAATAACCTCAGAAAGATGGCAGAAAGTGGTCATT 360
QY |||||
361 ATCAAGAAATGAGGTCAAGCTCAAGAGTTAAAAAATTGAGGGAAGATTGTTCCGCCCTCG 420
DB |||||
361 ATCAAGAAATGAGGTCAAGCTCAAGAGTTAAAAAATTGAGGGAAGATTGTTCCGCCCTCG 420
QY |||||
421 GGGTATGATGCTTCTCCAACTCGAAACGGGATCATACGGGAACCTCTGAGAAC 480
DB |||||
421 GGGTATGATGCTTCTCCAACTCGAAACGGGATCATACGGGAACCTCTGAGAAC 480
QY |||||
481 CAGCACCTAAAAACGAGGATCCACCCGAACTCCACCGATGTTCCCGGTTTATTG 540
DB |||||
481 CAGCACCTAAAAACGAGGATCCACCCGAACTCCACCGATGTTCCCGGTTTATTG 540
QY |||||
541 ATCTTATATACGCTGGAACGCTGGCGCTTCTACGGGAACGCGAGGACCGCA 600
DB |||||
541 ATCTTATATACGCTGGAACGCTGGCGCTTCTACGGGAACGCGAGGACCGCA 600
QY |||||
601 GGAACACCGGCAATACCGCGAATGCGCACGACCGTGTGTTGCCAAGCATGGTT 660
DB |||||
601 GGAACACCGGCAATACCGCGAATGCGCACGACCGTGTGTTGCCAAGCATGGTT 660
QY |||||
661 CGGCGCGGTGACGACCTGGGAGCGGAGTGAACACCTTATCCCTGTGTGAAGAG 720
DB |||||
661 CGGCGCGGTGACGACCTGGGAGCGGAGTGAACACCTTATCCCTGTGTGAAGAG 720
QY |||||
721 TCCTGCTGCGGCAATTCACCTCGAGGCGCTTTCATCAACGATCCCGTGTGTGCTC 780
DB |||||
721 TCCTGCTGCGGCAATTCACCTCGAGGCGCTTTCATCAACGATCCCGTGTGTGCTC 780
QY |||||
781 AAAACCGGATTTCAATTTTCCGGCAACCCACAGATCTTGCCCGGTGATCCATCGCG 840
DB |||||
781 AAAACCGGATTTCAATTTTCCGGCAACCCACAGATCTTGCCCGGTGATCCATCGCG 840
QY |||||
841 GAAAGGTTGATCAATTCATCAGTACGCGCGGAACCTGACATCTTCTGAGCTTC 900
DB |||||
841 GAAAGGTTGATCAATTCATCAGTACGCGCGGAACCTGACATCTTCTGAGCTTC 900
QY |||||
901 TCGATCTCTCGGAGCGCACCATCATCTTCTTCCCTCGGGCACACTGATGAGATTG 960
DB |||||
901 TCGATCTCTCGGAGCGCACCATCATCTTCTTCCCTCGGGCACACTGATGAGATTG 960
QY |||||
961 ATACCCTACAGGCAATTCCTTGGCTAAAGAGAAATGTAGCGGTACGGCTACGC 1020
DB |||||
961 ATACCCTACAGGCAATTCCTTGGCTAAAGAGAAATGTAGCGGTACGGCTACGC 1020
QY |||||
1021 ATTTGTTCAATGCGATCCCTCGGTGATCATAGGGCTCCCGCAGCGTGGGCTTTG 1080
DB |||||
1021 ATTTGTTCAATGCGATCCCTCGGTGATCATAGGGCTCCCGCAGCGTGGGCTTTG 1080

QY 1081 TTGCTGGGCACTGCCGGGACGCATATGTTGATGTTGATCGCGGCGTGCATTTGG 1140
DB |||||
1081 TTGCTGGGCACTGCCGGGACGCATATGTTGATGTTGATCGCGGCGTGCATTTGG 1140
QY |||||
1141 CCGATGGAACGTCGATCTAGCTGTTCCACACGCTTTTTCATCACGACGCCATGG 1200
DB |||||
1141 CCGATGGAACGTCGATCTAGCTGTTCCACACGCTTTTTCATCACGACGCCATGG 1200
QY |||||
1201 AAGCGCGGGAATCCAGACGCTGAGTACATTTTGGCGTTTGAAGCTCACCGTCAAG 1260
DB |||||
1201 AAGCGCGGGAATCCAGACGCTGAGTACATTTTGGCGTTTGAAGCTCACCGTCAAG 1260
QY |||||
1261 ATGAGTCCGCGCTGCGCGATGCGGCGCATCGCGGGGACAGACACATACGCA 1320
DB |||||
1261 ATGAGTCCGCGCTGCGCGATGCGGCGCATCGCGGGGACAGACACATACGCA 1320
QY |||||
1321 GTCAGTTCGTCACCACTGCGCAGGGTATGACGCTTATGAGCGGACCCCTCCACACT 1380
DB |||||
1321 GTCAGTTCGTCACCACTGCGCAGGGTATGACGCTTATGAGCGGACCCCTCCACACT 1380
QY |||||
1381 CAACCGTCGCGCTAAATTTCTCGGCTTGGGATCAGGAATCGCTAAATCCAACTCG 1440
DB |||||
1381 CAACCGTCGCGCTAAATTTCTCGGCTTGGGATCAGGAATCGCTAAATCCAACTCG 1440
QY |||||
1441 CAAATTTTGTGCTTTGACTCAACGCGCAGGTCGAAAGGTCATTTAGTGCATCAAG 1500
DB |||||
1441 CAAATTTTGTGCTTTGACTCAACGCGCAGGTCGAAAGGTCATTTAGTGCATCAAG 1500
QY |||||
1501 TACITTAAGTAGGATAAATCTATCTGATTTTAAAGAGTCCACCATGGAATCACTA 1560
DB |||||
1501 TACITTAAGTAGGATAAATCTATCTGATTTTAAAGAGTCCACCATGGAATCACTA 1560
QY |||||
1561 TCTGAAAGAGGAGCAAGAGTCCGCAAGAGTGTGAGTCTTAAATCGACCCCTCGCCA 1620
DB |||||
1561 TCTGAAAGAGGAGCAAGAGTCCGCAAGAGTGTGAGTCTTAAATCGACCCCTCGCCA 1620
QY |||||
1621 ACNAGGTTGGAACCTTGGGCTTGCACAGCATCTCACCACCTGAGTACCTACCAAGGC 1680
DB |||||
1621 ACNAGGTTGGAACCTTGGGCTTGCACAGCATCTCACCACCTGAGTACCTACCAAGGC 1680
QY |||||
1681 TCATTCGATGATGAAGCTGGGAGTGTCTATTCAAGAACTGCAAGGATCTTGTGTTG 1740
DB |||||
1681 TCATTCGATGATGAAGCTGGGAGTGTCTATTCAAGAACTGCAAGGATCTTGTGTTG 1740
QY |||||
1741 ATGATAGCTGGGACTAACCGTGCAGTGAACAGCTACTTTAAACACCTGCGCAAG 1800
DB |||||
1741 ATGATAGCTGGGACTAACCGTGCAGTGAACAGCTACTTTAAACACCTGCGCAAG 1800
QY |||||
1801 AGTTCACTGACACATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTCAAC 1860
DB |||||
1801 AGTTCACTGACACATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTCAAC 1860
QY |||||
1861 CTGATCCATAGCAACGCTGCGAGTATGAGGCAAGATCGCTGACAGATCCGTTGAG 1920
DB |||||
1861 CTGATCCATAGCAACGCTGCGAGTATGAGGCAAGATCGCTGACAGATCCGTTGAG 1920
QY |||||
1921 TTCATATCTTGGCATCGGGAACGCGCATCGCTTTTCAATGAACCATCATCTCTCT 1980
DB |||||
1921 TTCATATCTTGGCATCGGGAACGCGCATCGCTTTTCAATGAACCATCATCTCTCT 1980
QY |||||
1981 GTCAGACTGCAAAAGTCCAGGCGCTCACCCCTCAACCCGCTACCCAGGTTTGGGCACTTTC 2040
DB |||||
1981 GTCAGACTGCAAAAGTCCAGGCGCTCACCCCTCAACCCGCTACCCAGGTTTGGGCACTTTC 2040
QY |||||
2041 CTTCAACCATCGGAAGGTCCTCAACCCGCTACCCAGGTTTGGGCACTTTC 2100
DB |||||
2041 CTTCAACCATCGGAAGGTCCTCAACCCGCTACCCAGGTTTGGGCACTTTC 2100
QY |||||
2101 CCGCGCGGAAACATCTGTTGGTGAACCTGTTGAAGGAAAGCGGACATCCGCGG 2160
DB |||||
2101 CCGCGCGGAAACATCTGTTGGTGAACCTGTTGAAGGAAAGCGGACATCCGCGG 2160
QY |||||
2161 AACTGTGGAAGCCCACTGACTGCTTCTTGGCCAGGTTCCATCTCTGTAGTGCACAACT 2220

[illegible]

RESULT. T 2.

```

US-08-961-527-201/C
: Sequence 201, Application US/08961527
: Patent No. 5420135
: GENERAL INFORMATION:
: APPLICANT: Charles Kunsch
: TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
: NUMBER OF SEQUENCES: 391
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,527
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 201:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3895 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
US-08-961-527-201

```

Query Match 2.9%; Score 172.2; DB 4; Length 3895;
Best Local Similarity 46.5%; Pred. No. 1.4e-43;
Matches 875; Conservative 0; Mismatches 963; Indels 45;

3797	QY	GGCGAACGCATCCTCGGGACACATTTGGCGGGAAGACACAACTTGTGCGCCGCCACCACTGT	3856
3874	Db	GCCTTCTGCTTTTCTGAGGCAATCGGAGGCAAGAAATGTGACTGCGCTAACTCACTGT	3815
3857	QY	GCAAGCGCTTTACGGCTCGTGTCTAAAGACACCAAGAGTGTGGATCGCCAAAAGTCTGGAT	3916
3814	Db	GGACACGGATCGGCTTTCTTTTAGAGATGATAAGAAGCTAAATGTTTAAAGCTATCGAG	3755
3917	QY	GATGATCCAGATCTGNAAGGCACCTTTGAAACTGGCGGCATGTTCCAGATCATCGTCGGG	3976
3754	Db	TCAATTCACAGCTGTTAAAGGAACCTTTTCAAAATGCAGGTCAATTTACAGTAAATCAT	3699
3977	QY	CCAGGCGATGTGGATCATGTTTTCAAAGAAGCTCGATGACGCAACCTTCCAAAGACATCGCT	4036
3698	Db	--TGGAAATGACGTGCCCATCTTTTATATGATTTACAGCCGTTTCAGGTATTGAGGT	3641
4037	QY	GTGTCCACAGACAGCTCAAGAGATGTTGGCTTACACAGCCCAACTGGTTCAGCCGTGT	4096
3640	Db	GTITTCCAAAGAAGACGCCAAGTCTGCAGCTAAGAGTAATCAAAACGTGTGTCCAAGGTGT	3581
4097	QY	GTGAAGTATTGGCGGACATTTTCGCCGCTGATTCCAAATCTTGTTGGTGGCGGTCTG	4156
3580	Db	ATGACCACCTCGCGGAGATTTTACTCCGATTTATCCAGCCTTGATAGTCGAGGATTG	3521
4157	QY	CTCATGGCTATCAACAATGTGTTGGTTGGCAGGAT-----CTGTTGGTCCGCAATCA	4210
3520	Db	ATCCTCGGTTCGGTAATGCTGTGAAGGTGTCATTTGTCGATGTTGGATGCGCAAGCC	3461
4211	QY	CTGGTGGAGATGTTCCCTCAGATCAGCGGTGTTCTGAGATGATCAACCTGATGGCATCT	4270
3460	Db	ATCACAGATCCCTCTCAGTTTGGCGAGGTGTCATCACTCTCCTCTGGTTGCCTGGTGA	3401
4271	QY	GGCCGTTCCGCTTCTTCCAGTGTGTTGGTTTTCACGCGCAACCAAGCGTTTCGCTGGC	4330
3400	Db	GCTATCTTCAGTTTCTACAGCAGGATACATTGTGCTGTTCTCGTAAGATGGGAACC	3341
4331	QY	AATGAGTTCCTGGCGCGCGCATTTGATGGCATGCTGTTCACCAACCTGGTTAAACGGC	4390
3340	Db	AGCCAAATTTGGAAATGTTCTCGGAATCTGTTTGGTATCGCTCAGTTGCTCAATGCC	3281
4391	QY	TAGACGTGGCCGCCACCATGACCGCGGGGGAATGCCAA-----TGTTGGTCCCTGTTT	4444
3280	Db	TATGCGGTTGCTTCAACGCCACAGCATGATATCGCGCAAACTGGGTTTGGAAATTTGGC	3221
4445	QY	GGTTTGGATTTGCTCTCAAGCTGGTTACCAGGCGACCGTGTCTCCTGTGCTGGTGGTCTCT	4504
3220	Db	TATTTTACTGTTAATCGTATCGTTACCAAGCCCAAGTTATCCAGCCTGTGTTGCAGGT	3161
4505	QY	TGGATTTGGCAGGATCGAAGTTCCTGCACAAGCGACTCATGGGCACTCAGACATTC	4564
3160	Db	TTGAGTCTGCTTATCTTGAATCTTCTGCACAAGCATATCCCAAGATCAATTCTATG	3101
4565	QY	CTGATCACCCAGTGTGACTCTGCTGCTCACCGGTTTCCTTACGTTCAATTGCTATTGGT	4624
3100	Db	ATTTTTGTACCTTTCTTGTCATTGATCCAGCCTTGATTTTGGCTCATACTGTTTGGGA	3041
4625	QY	CCAGCAATGGCTTGGGTGGTACTTGTCTGGCACAGGT---CTGACAGGACTCTATGAT	4681
3040	Db	CCAATCGGTTGGACAATTGGACAAGGACTTTCATCAGTTGCTGTGGCAGGTTTAACTGGT	2981
4682	QY	TTCCGTTGGTCCAGTCGCGGCTCTGCTTTTCCGTTCTGGTCTACTACCAATCGTTATCACT	4741
2980	Db	CCAGTTAAATGGCTCTTCGTTGCAATTTTTTGGCGCCCTCTACGCTCCATTTGTCATCACA	2921
4742	QY	GGTCTGCACACGCTCTTCCGCCCAATTGAGCTGGAGCTGTT---CAACACAGGTGATCC	4798
2920	Db	GGTCTGCACCATATGACCAATGCCATTGATATACAAATGATGGGATGCTGTGGGCACCT	2861
4799	QY	TTCATCTTCGCAACCGCATCCATGGGCAATATCCGACGGGTGCAGCATGTTTGGCAGTG	4858

Db	2860	GCCTATGGCCAATGATTGCTCTTCTTAATATTGCTCAAGGCTCAGCCGTGTTGCGTAT	2801
Qy	4859	TTCTTCTTACGGAAGAGTGAAGAGCTCAAGGGCC---TTCCAGGTGCTTCAGGTGCTCC	4915
Db	2800	TATTTATGATCGCCATGATGAGCGTGAAGCTCAGGTTTCACCTTCTGCTCAAGCAATTCA	2741
Qy	4916	GCCTGTTCTTGGTATTACAGAGCTGGGATCTTCGGTGTGAACCTTCGCCCTGCGTGGCG	4975
Db	2740	GCCTATCTCGGTGTTACGAACAGCAGCTCTTTTGGGGTTAAGCTAAATATATTATCCA	2681
Qy	4976	TTCTACATTGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATGCACTCTTTGATATC	5035
Db	2680	TTTGTTGCTGGGATGACTGGTTACGCCCTTCAGCGCATGTATCCGTTACTTTAATGTA	2621
Qy	5036	AAGGCAGTTGCGTTGGCGCTGCAGGTTTCTTGGGTCTTCTTCTATTGATGCTCCAGAT	5095
Db	2620	ACTGCGGCTTCTATTTGATATCGGTGGTTTCCAGGATATCTCTCTATCAACCTCAATAC	2561
Qy	5096	ATGTCATGTTCTTGGTTTCGGGGTAGTTACCTTTTGTATCGCATTCGCGCAGCGATT	5155
Db	2560	ATGCTGCCATTTCAGGAACATGCTAGTTGCGATTGTTGCCAATGCTC-----TTG	2507
Qy	5156	GCTTATGGCTTTTACTTGTGTCGGCGCAAGCGGAGCATGATCCAGATGCAACCGCTGCT	5215
Db	2506	ACTTCTCTTCGCAAGGCTGGTCTCTTTACAAAAACAGAGGCGCATGCAACTTCGAG	2447
Qy	5216	CCAGTGCTCGAGGAAGCAGCAAGCGGAAGCAGAGCACCGCAGAGATTTTCAAACGAT	52757575
Db	2446	GCAGATTCTGTTGCTCAAGAGAGCAGAAATTTGTGAACCATGAACAGTAGAATCTACT	2387
Qy	5276	TCCACCATATCCAGGCACCTTTGACCGGTGAAGCTATCGCACTGAGCAGCGTCAGCGAT	5335
Db	2386	TCGSGTAAATATTACGCCCACTAACTGGCCACAGTGAAGAATTTGATCAAGCGACGGAT	2327
Qy	5336	GCCATGTTTGGCAGCGGAAGCTTGGCTCAGGTTTGGCATGTCGCCCCACCAAGGGCAG	5395
Db	2326	CCATATTTTGCATCAGGTGTCATGGGCAAGGCTACTCATTTGAACCAAGCAAGGTGAG	2267
Qy	5396	CTGTTTTCACAGTGAGCGGAAGATCGTGGTGGCTTCCCATCTGGTCAAGCTTTTCGA	5455
Db	2266	TTGACCTCTCCAGTTAATGGGACAGTACAGGTTCTTTTCCCTACCAAGCATGCCATCGC	2207
Qy	5456	GTCCGCACTAAGGCTGAGGATGGTTCCAAATGTGGATATCTTGATGACATGTTGTTGAC	5515
Db	2206	ATTGCTCTGACGAGG-----GAGTTGAATTGCTCTCATCCACATCGGTATGGAT	2159
Qy	5516	ACCGTAACCTTCAACGGCACCGCACTTTAACCCGTGAAGAGCAGGCGGATGAAGTCAAA	5575
Db	2158	ACAGTAGGCTTGTATGGCAAGGTTTGAAGTCTTGTAGTCCCAAGGAGATCAGGTTACA	2099
Qy	5576	GCAGGGAGCGTGTGTGAATTCGATTTGATGCCATTAAAGGCTGCAGGTTATGAGGTA	5635
Db	2098	GTTGGTCAGCAACTGATTCTGTTTGTATATGGATGTCAATTAAGGCTGCAGGTCGTGACA	2039
Qy	5636	ACCACGGCGATTGTTGTTCCAA	5658
Db	2038	GAAACTCTGTTATCATCACCAA	2016

RESULTS

US-08-961-527-30/c

US-08-301,527, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 110

: GENERAL INFORMATION:

APPLICANT: Charles Kunsch

: TITLE OF INVENTION: Stre

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
ADDRESS: Human Genome Sciences, Inc.

STREET: 2410 KEY WEST AVENUE
CITY: Rockville

CITY: ROCKVILLE
STATE: Maryland

STATE: MARYLAND

COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,527
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 30:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9769 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-527-30

Query Match 2.6%; Score 154.2; DB 4; Length 9769;
 Best Local Similarity 46.7%; Pred. No. 1.7e-37;
 Matches 917; Conservative 0; Mismatches 963; Indels 85; Gaps 10;

3779 ATGGACCATAGGACCTCGGCGACGATCCTCGGACATTTGGCGGACGATTTGGCGGCGAGACACAT 3838
 9710 ATGAACATCAGGAATTCGAAAGATCATCGATCGCTTGGCGGCGACGCTGAAATGTC 9651
 3839 GTCCGCCCGCACACTGTCAACGCGTTTACGCTCGTGTCTGCTCAAGACACACAGAGTGTG 3898
 9650 AATAAGTGTGCCACTGTGGGACTCGTCTACGTGTCATGTCGTAAGATGAAGAAATC 9591
 3899 GATCGCAAGATCTGGATGATGATCCAGATCTGAAGGACACCTTTGAACTGGCGCATG 3958
 9590 AATAAAGTGTGATGAGAACTTGGAAAGTTCAAGGTGCTTTCTTAAGTCAAGGCAA 9531
 3959 TTCAGATCATCGTTCGGCGGCGAGCGATGTGGATCATGTTTCAAGAACTCGATGACGCA 4018
 9530 TACCAATATCTTTGGTACAGGTACAGTTAAACAAATGTACGATGAAGTTGTGTACTT 9471
 4019 ACCTCCAAAGACATCGTGTGTCCACAGACAGCTCAAGATGTTGTGGTACACAGCC 4078
 9470 GGATTAACAA-----CATCATCTAAGGATGACATGAAGCAGAAAGTTGCTAAACAGGG 9417
 4079 AACTGGTTCAGCCGTGTGAAGGTATTGGCGGACATTTTCGTCGCGCTGATTCCTCAATC 4138
 9416 AACTGGTTCACCTGTATCCGTAATTTGGTGTATGTTTTCGTTCCAAATCATCCAGTT 9357
 4139 TTGTTGGTGGCGGTCTGCTCATGGCTATCAACAAATGTGGTGGCGAGATCTGTTTC 4198
 9356 ATCTAGCAGCAGCTCTCTCATGGGTGCGGTGCTTTTCAACGCTCTTGAATGCCA 9297
 4199 GGTCCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTGTGTGAGATGATCAAC 4258
 9296 CTTCAG-----GTACATTGCACTTACACCAA 9267
 4259 CTGATGATGATCTGGCGGTGTGGCTTCTTGCCAGTGTGGTGGTTTTCACCGCAACCAAG 4318
 9266 ATCTTGACAGATACAGCCTTCATCATCTTGCCAGGTTTGTGTGGTCAACCTTCGCT 9207
 4319 CGTTTCGGTGCATGATGATCTCTGGCGCGGATTTGGTATGGCGATGGTGTCCCAACC 4378
 9206 GTATTGGTGAATCTCGCGTGGTATGCTTCTTGGTATGATGCTTGTCTCTGGCTCA 9147

QY 4379 CTGGTTAAAGGCTACGAGTGGCCGCCACCACCATGACCGCGGCGAAATCCCAATGTGGTCC 4438
 Db 9146 CTTCACAAAGCTTGGCGAGTTGC-----TCAAGGTGTGAAGTAACACGATGAAC 9096
 QY 4439 CTGGTTGGTGGATGTGTCTCAGCTGGTTACAGGCGACCGTCTCTCTGTCTGTCTGTG 4498
 Db 9095 TCTTTTGGTTT---CATCCCTGTGTGGTTGCAAGGTTCGCTTCTTCCAGCCTTCAIC 9039
 QY 4499 GTCTCTTGGATCTCTGGCAACGATCGAAGATTCCTGCAACAGGACTCATGGGCACTGCA 4558
 Db 9038 ATCGGGGTTGTGGAGCTAAATTTGAAAAAGCTGCCGCAAGTTGTTCCAGATGTC-ATT 8980
 QY 4559 GACTTCCTGATCACCCAGTGTGTACTCTGTCTGTCTCAGCGGCTCTCTTACCTTACCTTATGCT 4618
 Db 8979 GACTCTTGGTAACACCATTCGTGTACACTTTTGGTCAATGTATCTCTTGGACTCTTTGTC 8920
 QY 4619 ATTGGTCCAGCAATCGCTGGTGGTGTCTGTGGCACACGGTCTCTCAGGGAATCTAT 4678
 Db 8919 ATTGCAACAGTTTCCACGTTGTTGAAAACTACATCTTATGCTACAAAAAGCGATTCTT 8860
 QY 4679 GATTTGGTGGTCCAGTCCGGGCTGCTTTTCGGTCTGTCTACTACCAATCGTTATC 4738
 Db 8859 AGCATGCCATTTGGTCTTTGGTGTCTTGTGTTGGTGGGTTCACCAATGATCGCTGTG 8800
 QY 4739 ACTGTCTGCACCACTGCTTCCCGCAATTTGAGCTGGAGCTGTTCAACACAGG---TGA 4795
 Db 8799 TCAGGTGTGCACCACTCTTCACTGCTTGAAGTGCATTTGCTGTCTGACCATGCT 8740
 QY 4796 TCTTTCATCTTCGCAACCGCATCCATGGCCAAATATCGCGAGGCTGCGAGCTGTTGGCA 4855
 Db 8739 AACCCATTCAAGCTATCATCACAGCTGCTATGACAGCTCAAGTGTCTGCTACTGTCG 8680
 QY 4856 GTGTTCTTCTAGCGAAGAGTGAAGGCTCAAGGCTTCAAGGCTTGCAGGTGCTTCAGTGTCTCC 4915
 Db 8679 GTTGGTGTAAACAAAAAATCCAAACTGAAACACTGCTTCCCGGCTGCTCTTCT 8620
 QY 4916 GCTGTCTTGTGATACAGAGCTCGCATCTTGGTGTGAACTTTCGCTGCGCTGCGGTGGCG 4975
 Db 8619 GCTTCTCTAGTATTACAGAGCTGCTATCTTTCGGGTGAACTTTCGCTGCTGCTGTAACCA 8560
 QY 4976 TTTCTATTGTTGATCGGTACCGCAGTATCGGTGCGGCTTTCATTGCTCACTCTTTTATATC 5035
 Db 8559 TTTCTTCTTTCATTGATGCTGTGTCATCGGTGTGGTGTGCTTCTATCTCTTGGACTT 8500
 QY 5036 AAGGCAAGTTCGTTGGGCGCTCGAGTTTCTTGGGTGTGTTTCTTATTGATGCTCCAGAT 5095
 Db 8499 GCTGTGATCTGTAATGTATCACCATCATCCCTGGTACAAATGCTTATGTTGGTAACGGA 8440
 QY 5096 ATGGTCATCTTCTTGGTTTGGCGGTAGTACCTTTGTCATCGCATTCGGCGCAGCAT 5155
 Db 8439 CAACTTCCAAATACCTTCTTATGTTAGTGTAT-----CATTTGCCCTTGGTTT 8389
 QY 5156 GCTTATGGCTTTTACTTGGTTCGCGCAACGCGAGCATTTGATCCAGATGCAACCGCTGCT 5215
 Db 8388 GCTCTTACTTACATGTTTGGTTACGAAG-----TGAAGTAGACGCAACTGCAGT 8338
 QY 5216 CCAGTCCCTTCAGGAACGACCAAGCCGAGCAGACAGCAGACCGAGATTTTCAACAGAT 5275
 Db 8337 GCAAAACGAGCTGAAGTGTCTGAAGAAAGAGAGAGTTGCCCGACGCTCTTCAAAAT 8278
 QY 5276 TCCACCATCATCCAGGCACTTTGACCGGTGAAGTATCCGACTGAGCAGCTGAGCGAT 5335
 Db 8277 GAAACACTGTGA---ACTCTATCTGCTGCTGATGTTGCTGCTTGTGATGCTCAATGAC 8221
 QY 5336 GGCATGTTTGGCAGCGGAAAGCTTGGCTCAGGTGTTCCGATCGTCCCCACAGGGCAG 5395
 Db 8220 CCAGTCTTCTCAAGTGGAGCTATGGGACAAGTGTGTTGTAACCAAGCAAGCGCGTG 8161
 QY 5396 CTGGTTTCCAGTGAAGGAAAGATCGGTGGGCTTCCCATCTGGTCACGCTTTCGCA 5455
 Db 8160 GTCTATGACACAGCTGATGCTGAAGTTTCAATTGCTTCCACAGGGCAGCCTTTGGT 8101
 QY 5456 GTCCGCACTAAGCTGAGGATGTTTCCAAATGTGGATGCTTCTGATGACATTTGGTTTCGAC 5515

Db 8100 TTGAAACAGAAATG-----GTGCTGAAGTTTGTATCATCTTGGTATTTGAT 8053
QY 5516 ACCGTAACCTCAACGGCACGCACTTTAACCCGCTGAAGAAGCAGGGCGATGAAGTCAAA 5575
Db 8052 ACTGTATCATGAACGGTACGGTTTGAACAAAAAGTTGCTCAAGGTAATAAGGTGAAA 7993
QY 5576 GCAGGGAGCTGCTGTGTAATTCGATATGATGCAATTAAGCGTGCAGGTTATGAGTA 5635
Db 7992 GCTGGCGATGTTCTTGAACATTTGACTCAAAACAAATCGCTGCAGCTGGAATGATGAT 7933
QY 5636 ACCACCCGATTTGTTTTCGAATTAACAAGAAACCGGACCTGTAAACACACTTACCGTTTG 5695
Db 7932 ACAACAATGTTATCGTTTCAAAATACAGGTGACTAGCTTCAAGTCCAGTCGCAACA 7873
QY 5696 GCGCAATTTGAAGCGGAGCCACCTGCTCAACGTCGCGAAAGAAA 5740
Db 7872 GGTTCAGTTGTAAGGGGATGCTGTGATGCAAGTGAAGTAAATCTAA 7828

RESULT 4
US-09-134-001C-1277
; Sequence 1277, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1277
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (20),(22),(32)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.

US-09-134-001C-1277

Query Match 2.3%; Score 139; DB 4; Length 732;
Best Local Similarity 56.6%; Pred. No. 1.5e-33;
Matches 306; Conservative 0; Mismatches 220; Indels 15; Gaps 2;

QY 4569 TCACCCAGTGTGACTCTGCTGCACCGGCTTCTTACGTTCAATTTGTTATGTCGCGAG 4628
Db 80 TGNCGCATATATCAATTTTATACAGCAATTAACATTTTATTTGAGGCGCTG 139
QY 4629 CAATGCGTGGGTGCTGCTGCTGGCACACGGTCTCGAGGACTCTATGATTTGCGGTG 4688
Db 140 TCACTCGTCAATAGTGTATTTGTTATCTGATGATGATGCTTGGTTATGATGAATTTGGTG 199
QY 4689 GTCCAGTGGCGTCTGCTTTTGGTCTGCTACTCAACATGTTATCACTGCTCTGC 4748
Db 200 GAGCTATGTTGACTTATTTGTTTATATGCGCAATCGTCAATCAGGAATGC 259
QY 4749 ACCAGTCTTCCCGCCCAATGAGCTG-----CAGCTGTTCAACAGGTTGAT 4796
Db 260 ACCATATCTTTATGAGTTGAACAGACATTAATTTGCTGATCGACTAAACAGTGGTT 319
QY 4797 CTTTCATCTCGCAACCGCATCCAGGCAATATCGCGAGGTCGAGCATCTTTGGCAG 4856
Db 320 CATTTATCTCCCAATCGCAACGATGCTCAAAATTTGCAACAGGTGGTGCAGGTTAGTGTG 379
QY 4857 TGTTCCT---TCCTAGCGAAGTGAAGAACTCAAGGCGCTTGCAGGTGCTTCAGGTGCT 4913
Db 380 CATTCCTTATCATATAGCAAAATATAAAATTAAGGTGTTGCTCCGCGCGGTATTT 439

QY 4914 CCGCTGTTCTTGGTATTATACAGAGCCTCGGATCTTCGGTGTGAACCTTCGGCTCGGCTGGC 4973
Db 440 CAGCTTTACTAGGAATTTACAGAACCAATGTTTGGTGTCAATCTTTAAATTTGAGATATC 499
QY 4974 CGTTCTACATTTGGTATCGGTACCGCAGCTATCGGTGGCGCTTTGATTGCACTCTTTGATA 5033
Db 500 CATTTATAGGTGCTGTTGCAGGATCAGGTATAGGTGGCGCTTATATTTCAATCTTCAAG 559
QY 5034 TCAAGCAGTTGCTGTTGGGCGCTGCAGGTTCTTFFGGTGTGTTTCTTATGATGCTCCAG 5093
Db 560 TAAAGCGATAGCGCTTGGTACAGCTGATACCTGATTTATATCTATAATATCTACAC 619
QY 5094 A 5094
Db 620 A 620

RESULT 5
US-09-107-532A-378
; Sequence 378, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 378:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8) LOCATION 1...714
; SEQUENCE DESCRIPTION: SEQ ID NO: 378:
US-09-107-532A-378

Query Match 2.1%; Score 124.8; DB 4; Length 714;
Best Local Similarity 52.5%; Pred. No. 5e-29;

	Matches	380;	Conservative	0;	Mismatches	322;	Indels	22;	Gaps	4;
QY	1544	CACCATGGAATC	AATCACTATCT	CGAAAGACGACG	AGAAGTCGG--	-CAAAGCAGTTCG	AGT	1600		
Db	9	CACATGGAATC	ATCATTCGT	TAAAAAACGC	TGAAGAAAGCG	GAAAAAAGCATTT	GAAT	68		
QY	1601	CCTAATCGCAC	CCCTTCGCCAAC	ACAGGGTGGA	ACCCTTGGGCTT	GGCAACGAGATCCT	CAAC	1660		
Db	69	CATCAAGAAG	GTATGAACAAG	GGTCCAAAGGT	TATTAGGATTGG	CAACGAGGACAC	CAAC	128		
QY	1661	ACTGAGTAGCT	ACCAAGAGCT	ATTCGCATGT	ATGAAGCTGGG	GAAGTGTCTAT	CAAGAA	1720		
Db	129	AGTTACTTT	TATAAAGAAT	TAAGTCTCAA-----	-----	CGCATTTGG	ATTCTAGTGA	176		
QY	1721	CTGCAAGGC	ATCTTTGGATGA	ATACGTGGGACT	TAACCCGTGAC	GTACGATGAAA	ACAGCTA	1780		
Db	177	AATGACTTCT	GTAAACTTAG	ATGAGTACGT	TGGACTAGCGGT	TTCAGATGAACA	AAAGCTA	236		
QY	1781	CTTTAAACCA	TCGCAAGAGT	TTCACTGNACC	ACATCGACATCG	TTCGTTGATGA	GAGAGCTTA	1840		
Db	237	CGGTCAATTC	ATCATGGATG	TCCAAATATT	TGTATAAAAA	ACCATTCA-----	AAGAAACGTT	290		
QY	1841	CAGCCAGATG	TGTGTAACCC	CTCATACGAAG	CAGCTGTCAGAGT	TATGAGGCA	AAAGAT	1900		
Db	291	TGTACCAGAT	GTAAGACAG	AGATCTGGNAG	CAGCATGCCAAA	AGTATGATCA	ATCAT	350		
QY	1901	CGGTGCGAAT	CCGTTGAAAGT	TCAAATPCCT	TGGCATCGGCGG	AAACGG-CAC	ATCGCTTT	1959		
Db	351	CGATGCACAT	CCCAATCGAT	ATCCAATCTTAG	GAATCGGGCAAA	TATGGACATAT	TCGTTTT	410		
QY	1960	CATTGRAACCA	TATCTTCCTG	TCAGGACTGCA	AAAGGTCCAGCGCT	GCACCC	TAAAC	2019		
Db	411	CAATGAACCT	GTGGTCTTAT	TTCAGCGAAG	CAGACTTCAGTAG	TTGACTTAACA	GAATCAAC	470		
QY	2020	TGTGGAGGACA	ACGCTCGATTCT	TCAACCCATCG	GAAGGTCCCAC	ACCACCGCGT	CAAC	2079		
Db	471	AATCAATGCA	ACAAACAGGAT	CTTCGNTAAG	CAGAGATGTAC	CAACACAAA	AGCTGTTT	530		
QY	2080	CCAGGTTTGGG	CACATTTGTC	CGCGCGCAAA	ACATCGTGTGGT	GGCACTCGT	GAAGG	2139		
Db	531	AATGGGAAT	CGGTTCAATT	TATGAAAAAG	TAAAGAAATCGT	CTTCA	TGGCATTTGGCG	AAAG	590	
QY	2140	AAAGCGGAC	GCATCGCGG	AACTGTGGG	BAAGCCCCAGT	GCATCTCTT	TGCCCAGGTT	2199		
Db	591	CNAAGTGTG	CAATCAAAAGG	ATGATTTGAAG	GACCTGTAA	CGAAATCATTTG	CCGACCAAG	650		
QY	2200	CATTCCTGT	AGATGCACAA	CAATGCCAC	CATCATCGTTGG	ATGTGAAG	CAGCAGTATCC	AAAGCT	2259	
Db	651	TGTATTGCA	AAATCAGC	ACAAATGTT	GTGTGATCAT	CGATGAGAG	CACCAAGT	710		
QY	2260	GGAA	2263							
Db	711	GTA	714							

RESULT 6

US-08-920-812-17/c
; Sequence 17, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rio-Laures Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 3615 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: Clinical isolate EC-24
US-08-920-812-17

Query Match.	1.8%;	Score 108.6;	DB 1;	Length 3615;
Best Local Similarity	45.5%;	Pred. No. 2.9e-23;		
Matches 666;	Conservative	0;	Mismatches 754;	Indels 45; Gaps 6;
QY	4270	TGCGCGTTCGCGCTTTGGCAGCTGTTGGTTGGTTTCACCGCAACCAAGCGTTTCGGTGG	4329	
Db				
Db	2786	TGCGTGGTTTGGTTCTTCCGATATCTCGGATACACCGGGGAAACGCTTCGGCGA	2727	
QY	4330	CAATGAGTTCTTGGGCGCGGCATGGATGGCGATGGTGTTCCTCCAAACCCCTGGTTAACGG	4389	
Db				
Db	2726	TAATCCATTACTGCCATTGGTGATTTGGTGAGCGGTTAGTCGATCCCAATTAATTCTGTACTG	2666	
QY	4390	CTAGGACGTGGCGCGCCACCATGACCGCGGGCGGAATGCCAATGTGTGTCCTGTTGGTTT	4449	
Db				
Db	2666	CTTTTCGAGAACGCGCAAAA-----GCGGATGCGCTGGGGCTGGATTCTCTGGGTAT	2615	
QY	4450	GGATGTTGTCFAAGCTGGTTTACCAGGCGACCGTGTCTCTGTCTGGTGGTCTCTTTGGAT	4509	
Db				
Db	2614	TCGGTTCACATTGTTGNAITACTTCGTCATCGGTTATCCCATTAATTTTTTCGCGCTGTT	2555	
QY	4510	TCGCGCAACGATCGAGAAGTTCTCTGCACAAGCAGCTCATGGGCACCTGCAGACTTCTGAT	4569	
Db				
Db	2554	GTGCAGCATTCTGGAACCGCGACTTAATGGTGGTTACCGTCGGCAATCAAAAATTTCTT	2495	
QY	4570	CACCCAGTGTGACTCTGTCTGCTCACCCGGCTTCTTACGTTCATTGCTATTTGGTCCAGC	4629	
Db				
Db	2494	CACACCAATGCTATGCTGATGTTATFACACCCGTCACCTTTCTGCTGGTGGGGCGCT	2435	
QY	4630	AATCGCTGGTGGGTGACTTGTGTCACACGCTCTGCAGGGACTCTATGATTTCGGTGG	4689	
Db				
Db	2434	ATCAACCTGGATAAGCGAATGATTGCCGCGGTTATCTCTGGCTTTATCAGSCGGTTCC	2375	
QY	4690	TCAGTCGGGGGTCTGCTTTTTCGGTCTGGTCTACTCACCAATCGTTATCACGTGGTCTGCA	4749	
Db				
Db	2374	TGCAATTCGCGGGCGGTAATGGCGGGCTTCTGGCAAAATCTTCGTGATCTCGGACTGCA	2315	
QY	4750	CCAGTCTCTCCCGCCAAATTCAGCTGGAGCTGTTCACCGAGGTGGATCTTCACTCTCGC	4809	
Db				
Db	2314	CTGGGGCTCTGGTCCGGGTGTATCAATAACTTTCACCGTGTCTGGGTACGACCAATGAT	2255	
QY	4810	AACCGCATCATG---GCCAATATCGCGCAGGTGGCATGTTTGGCAGTGTCTTCCT	4866	

2254 CCCGCTGTTAATGCCCGCCATTATGCGCAGGTGCGGGGCGCTCGGCGTCTTCTCTG 2195
 QY 4867 AGCGAAGATGAAGAGCTCAAGGCGCTTCAGAGTGTCTCAGTGTCTCGCTGTTCTTGG 4926
 Db 2194 CGAACCGGATGCGCAGAAAAAGTGGTGGCGGATGACGGCGTTGACGAGTCTGTTGG 2135
 QY 4927 TATTACAGAGCCGCGCATCTCGGTGTGAACCTGCGCTGCGTGGCGCTTCTACATGG 4986
 Db 2134 TATCACCAGACAGCGGTATGCGGTCAACCTGCGGTAAGTACCCCTTGTATCGC 2075
 QY 4987 TATCGTACCGAGCTATCGGTGGCGCTTGTATGCACTCTTGTATCAAGCAGTTCG 5046
 Db 2074 CTGTATCAGTGGGCTTGGGGCCACCATTTATGCTACGGGCAAGAAAGTCTACTC 2015
 QY 5047 GTTGGCGGTGCGAGGTTCTTGGGTGTGTTTATGATGCTCCAGATATGTCATGT 5106
 Db 2014 CTTTGGTTGCCAAGTATTTTACCTTCATGCAACCATCCCGTCAACGGGAATGATTT 1955
 QY 5107 CTTGGTTTGGCGGTAGTTACCTTTGTCTATCGCATTCGGCGAGCGATTTATGGCCT 5166
 Db 1954 CACCGTCTGGCGCAGCGTTA---TTGGCGGTGTCTATGCGCATGGTGGCAATTTGTCGG 1898
 QY 5167 TTACTTGGTTCGCGCAAGCGCAGCATTTGATCCAGATGCAACCGCTGCCAGTGCCTG 5226
 Db 1897 TACGGTGTGCTTCATTTTCATCCCGTAAACGCTCAGCCAGC-----1856
 QY 5227 AGAAGCAACAAAGCGAAGCAGAGCAGCAGCCCGCAGAAATTTCAACAGATTCACCATAT 5286
 Db 1855 -GCAGGTGCGCGAAGAAACACAGAGGTTATTACACCACTGAGCAGGCGGTAT 1797
 QY 5287 CCAGGACCTTCCAGCGTGAAGCTATCGCAGTGCAGCGTGCAGGATGCGATTTTCG 5346
 Db 1796 CTGTTCCAGATCAGCGGAGAGATTGTGCGCTCATCTCAGTGCCTGATACACGTTGC 1737
 QY 5347 CAGCGAAGCTTGGCTCAGTGTGCGATTCGATCGTCCCAACAAAGGCGAGCTGTTTACC 5406
 Db 1736 CAGTGGCTGTGGTAAAGTATTGCCATTCTGCCCTCGTGGTGAAGTGGGTTCTCG 1677
 QY 5407 AGTACGCGAAGATCGTGGCTTCCCATCTGCTCAGCTTTCGACGTCGCGACTAA 5466
 Db 1676 GGTGCGGGTGAATTCCTGCTGTTGCGGCATTTACGCGCATG-----GCATGTA 1623
 QY 5467 GGTGAGATGGTTCCAAATGGATATCTGATGCACTTGTGATGCACTGTTTCGACACCGTAAACCT 5526
 Db 1622 GTCAGATGATG-----TGTGGAGATCCTGATTCATGTCGCTATCGACACCGTAAACT 1569
 QY 5527 CAACGCGACGACTTTAACCGCTGAGAGACGAGCGGATGAGTCAAAACGAGGAGCT 5586
 Db 1568 GGACGGCAATTCCTTCCGCTCAGTCAACGTTGGTGCAAGTCAATACAGCGCATG 1509
 QY 5587 GCTGTGAAATTCATATGATGCAATTAAGGCTGCGAGTTATGAGTAAACACCGCGAT 5646
 Db 1508 GCTGATTTCTTTGATATCCTGCTATTCGAGGCGGATTTGATCTGACGACGCGGT 1449
 QY 5647 TGTGTTTCGAATTACAGAAACCGACCTGTAACACATTCAGGTTTGGCGCAATTGA 5706
 Db 1448 ATTAATCAGTAATAGCATATTTTACGACGATTTATACCCACCGGACGCGGAGATAAG 1389
 QY 5707 AGCGGGAGCCAACTGCTCAACGTC 5731
 Db 1388 CGCAGGTGAACCGCTGTTATCCATC 1364

RESULT 7

us-08-920-827-17/c
 ; Sequence 17, Application US/08920827
 ; Patent No. 5770375
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsunisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji

TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 6300 Sears Tower, 233 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/920.827
 FILING DATE: 29-AUG-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/362,577
 FILING DATE: 27-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 19036/32420
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312/474-6300
 TELEFAX: 312/474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3615 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli
 STRAIN: Clinical Isolate EC-24
 US-08-920-827-17

Query Match 1.8%; Score 108.6; DB 1; Length 3615;

Best Local Similarity 45.5%; Pred. No. 2.9e-23;

Matches 666; Conservative 0; Mismatches 75; Indels 45; Gaps 6;

QY 4270 TCGCGCGTTCGGGTCTTCCAGTGTGTTGGTTTACCGCAACCAAGCGTTTCGGTGG 4329
 Db 2786 TCGCTTGTGTTTCCCGGATTAATCCTGGATACACCGGGGAAACGCTTCGGCGA 2727
 QY 4330 CAATGAGTTCCTGGGCGCGGCGGATGATGCGGATGTTCCCAACCCCTGGTTAACGG 4389
 Db 2726 TAATCCATTTACTGCGATGTTGTTGGAGCGTTAGTGATCCATTAATCTGTACTG 2667
 QY 4390 CTACGACGTGGCGCCACCATGACCGCGGGGGAATGCCAATGTGGTCCCTGTTGGTTT 4449
 Db 2666 CTTTTCGAGACCGCGCAAAA-----CGGATCGCTGGGCTGGATTTCTCGGTAT 2615
 QY 4450 GGATGTTGCTCAGCTGTTTACAGGCGACCGTCTCTGCTGCTGGTGGTCTCTGTTGAT 4509
 Db 2614 TCGGTGTCATGTTGAATTAATCTGCTGATCGGTATTCCTCATTTTTCGCTGCTGTT 2555
 QY 4510 TCTGGCAACGATCGAGAGTTCTCTGCAAGCGACTCATGGCGACTGCGACTTCTCTGAT 4569
 Db 2554 GTGCAGCATTCGGAACCGCGGACTTAATGCGTGGTTACGTCGGCAATCAAAATTTCTT 2495
 QY 4570 CACCCAGTGTGACTCTCTGCTCAGCGGCTTCTTACGTTCAATGCTATTTGGTCCAGC 4629
 Db 2494 CACACCATTTGCTATGCTGATGTTATCACACCGCTCACCTTCTGCTGGTGGGCGCCT 2435
 QY 4630 AATGCGCTGGTGGTGAATTTGCTGCAACGACTGCTGAGGAGACTCTATGATTTGGTGG 4689
 Db 2434 ATCAACCTGATAAGCGAACTGATTCGCCCGCTTATCTCTGCTTTATCAGCGCGTTC 2375

QY 4690 TCAGTCGGGGCTCTCTTTTGGTCTGTCTACTCAACAACTGTTATCACTGGTCTGCA 4749
Db 2374 TGCATTGGGGGGGGTAAATGGCGGCTTCTGGAAATCTTCGTCTATCTTCGGACTGCA 2315
QY 4750 CCAGTCTTCCCGCCCAATAGCTGGAGCTGTTCAACAGGGTGGATCTTTCATCTTCG 4809
Db 2314 CTGGGCGCTGGTSCCGGTGTATCAATAACTTCACTGCTGGGTACGACACCATGAT 2255
QY 4810 AACCGATCCATG---GCCAATATCCGCGAGGTGACGATGTTGGCAGTGTCTTCT 4866
Db 2254 CCGCTGTAAATCCCGCCATTAATGGCGAGGTGCGGGGGCGGCTCGGCTCTTCTCTG 2195
QY 4867 AGCGAAGTGAAGTCAAGGGCTTGCAGTGTCTCAGTGTCTCCGCTGTTCTTGG 4926
Db 2194 CGAACCGATGCGCAGAAAAAGTGTGGGGGATCAGCGGCTTGACGAGTCTGTTGG 2135
QY 4927 TATTACAGACCTGCGATCTTCGGTGTGAACCTTCGCTGCGGTGCGGCTTCTACATGG 4986
Db 2134 TATCCGAAACCGAGTATGCGCTCAACCTGCGCGCTAAGTACCCCTTGTATCG 2075
QY 4987 TATCGGTACCGCAGCTATCGTGGCGCTTTGATTGCACTCTTTGATATCAAGGCACTTC 5046
Db 2074 CTGTATCAGTGGGCTTGGGGCCACCATATTGGCTACGCGCAACGAAAGTCTACTC 2015
QY 5047 GTTGGCGCTGCAAGTGTCTGGGTGTTGTTCTATGATGCTCCAGATATGCTCATGTT 5106
Db 2014 CTTTGGTTTCCAAATTTTCACTTCATGCAAAACCATCCCTCAACGGGAATGATT 1955
QY 5107 CTTGGTTGGCGGTAGTTACCTTGTTCATCGCATTCGGCGAGCGATGCTTATGSCCT 5166
Db 1954 CACGCTGGCGCAGCGTTA---TTGGCGGTGTCATTCGATCGTGGTGGCATTTGCG 1898
QY 5167 TTACTTGGTTCGGCGCAACGGCAGCATGTATCCAGATGCAACCGCTGCTCCAGTGGCTGC 5226
Db 1897 TACGGTATGCTTCATTTCATCCGCTAAACGTCAGCCAGC----- 1856
QY 5227 AGAACCAACCAAGCCCAAGCAAGACCCGCGAGAAATTTCAACGATTCACCATCAT 5286
Db 1855 -GCAGGTGCCCCGAGAGAAAAACACAGAGGTATTACACCACTGAGCAGGGCGGTAT 1797
QY 5287 CCAGGCACCTTTACCGGCTGAAGCTATCGCACTGAGCAGCTCAGGCGATGCCATGTTGC 5346
Db 1796 CTGTTACCGATGACGGGAGAGATGTTGTCGCTCATTCAGCTGCTGATACCACTGTC 1737
QY 5347 CAGCGGAAGTGGCTCAGTGTGGGATCGTCCCGACCAAGGGGAGCTGTTTCACC 5406
Db 1736 CAGTGGCTGTGGGTAAAGTATTGCCATCTCGCCCTGCTGGTGGTGAAGTGGCTTCC 1677
QY 5407 AGTGAGCGGAAAGATCGGTGGCTTCCCATCTGTCAGCTTTTCAGTCCGCACTAA 5466
Db 1676 GGTTCGGGTGCAATTCCTGCTGTTGCGCACATTACAGCCATTG-----GCATTGA 1623
QY 5467 GGCTGAGATGGTTCGAATGGGATATCTTGATGCAATGTTTTCGACACCGTAAACCT 5526
Db 1622 GTCAGATGATGG-----TGTGGAGATCCCTGATTCATGTCGATCGGATCGACACCGTAAACT 1569
QY 5527 CAACGCGACGACTTTAAACCGCTGAAGAAGCAGGGGATGAAGTCAAAAGCAGGGAGCT 5586
Db 1568 GCAGCGCAATCTCTTCCGCTCAGCTCAAGTGGGTGACAAGGTCAATACAGCGATCG 1509
QY 5587 GCTGTGAATTCGATATGATGATTAAGCTGACAGGTATGAGGTATACACGCGCAT 5646
Db 1508 GCTGATTTCTTTTATATCCCTGATTCGCGAGGCGGATTTGATCTGACGACGCGGT 1449
QY 5647 TGTGTTTCGAATACAGAAACCGGACCTGTAAACACATTACGTTTGGCGCAATGA 5706
Db 1448 ATTAATCAGTAATAGCGATGATTTACGGACGTATTACCCACGCGCAGGCAAGTAAG 1389
QY 5707 AGCGGGACCAACCTGCTCAACGTC 5731
Db 1388 GCGAGGTGAACCGGTGTTATCCATC 1364

RESULT 8

US-08-921-177-17/C
; Sequence 17, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-24
; US-08-921-177-17
Query Match 1.8%; Score 108.6; DB 1; Length 3615;
Best Local Similarity 45.5%; Pred. No. 2.9e-23;
Matches 666; Conservative 0; Mismatches 754; Indels 45; Gaps 6;
QY 4270 TCGGCGGTTCGGGTTCTTCCAGTGTGTTTCCACCGCAACCAACCGTTTCGGTGG 4329
Db 2786 TGCCTGTGTTGTTCTTCCGATATCTCGGATACACCGCGGGAACCGTTCCGGGA 2727
QY 4330 CAATGATGTCCTGGGCGCGGATTCGTATGCGATGTTTCCCAACCCCTGTTAAACGG 4389
Db 2726 TATCCATTACTGCCATGTTGTTGTTGAGCGTTAGTCATTAATCTGTACTG 2667
QY 4390 CTACGACGTGGCGCCACCATGACCGCGGGAATGCCAATGTGTCCTCTTGGTGT 4449
Db 2666 CTTTTCGAGAACCGCAAAA-----GCGGATGCGTGGGCTGGATTCTCTGGGTAT 2615
QY 4450 GGATGTTGTCAGCTGGTTACAGGCGACCGTCTCTCTGCTGGTGTCTCTTGGAT 4509
Db 2614 TCGGTCACATGTTGAATTAATTCGTCATCGGTTATTCCTCATATTTTCTCGCTGTT 2555
QY 4510 TCTGGCAACGATCGAGAAGTTCTCTGCACAAGCGACTCATGGGCACTGCAGACTTCTGAT 4569
Db 2554 GTGCAGATTCTGGAACGCGGACTTAATGCTGTTACCGTGGCAATCAAAAATTTCTT 2495

QY	4330	CAATGAGTTCCTGGGCGCCGGCATTTGATATGCGCATGGTGTCTCCCAACCCCTGGTTAACGG	4389
Db	2726	TAAATCCATTTACTTGCCATGATGTAATGGTGAGCGGTAGTAGTCATCCCAATAATCTGTACTG	2667
QY	4390	CTACGAGTGGCGCCACCAATGACCGCGGCGCAAAATGCCAAATGGTGCTCCCTGTTGGTTT	4449
Db	2666	CTTTTCAGAAACGCGAAAAA-----GGGATGGCTGGGCTGGATTTCTCGGGTAT	2615
QY	4450	GGATGTGTCAAAGCTGGTTACCAAGGCAACCGTGCTTCGTGCTGCTGGTGTCTCTTGGAT	4509
Db	2614	TCCGGTCACAATGTTGAAATTAATTCATCGGTATATCCCAATTAATTTTCTGCCCTGGTT	2555
QY	4510	TCATGGCAACGATCGAAGATTCCTGTGCACAAAGCACTATGGGCACCTGCAGACTTCCTGAT	4569
Db	2554	GTGACGATTTCTGSAACGCGACTTAATGGTGCTTACCGTCGGCAATCAAAATATCTT	2495
QY	4570	CACCCAGTGTGACTCTGCTGTCTCACCGGCTTCCTTACGTTTCATTCATTCATTCGTCACG	4629
Db	2494	CACACCAATGCTATGCTGATGTTATCACACCCGCTCACCTTTCGTTGCTGGTGGGGCGCT	2433
QY	4630	AATGCGGTGGGTGGGTGACTTGTGTGGACACAGGCTGTCAGGGACTCTATGATTTCCGTGG	4689
Db	2434	ATCAACCTGATTAAGCAACTGATTTGCGCGCGTTATCTCTGGCTTTATCAGGCGGTTCC	2375
QY	4690	TCAGTCGGCGGCTGCTTTTCGCTGTGCTACTCACCAATCGTTATCACTAGTCTGCTGCA	4749
Db	2374	TGCATTTGCGGCGGCTAATGCGGCGCTTCTGCGCAATCTTGTCATGTTCCGACTGCA	2315
QY	4750	CCAGTCTCTCCCGCAATGAGCTGGAGCTGTTTCAACAGGGTGGATCCTTCATCTTCGC	4809
Db	2314	CTGGGCGCTGGTCCCGGTGTGTATCAATAACTTACCGTCTCGGGCTACGACACCATGAT	2255
QY	4810	AACCGCATCATG---GCCAATATCGCGAGGTGACAGATGTTTGGCAGTGCTTCTCCT	4866
Db	2254	CCGCTGTTAATGCCCGCATTAATGGCGAGTCTGGGCGGCGCTCGGCGTCTCCTCTG	2195
QY	4867	AGCGAAGATGAAGAAGCTCAAGGCGCTTGAGTGCTTCAGTGCTTCAGTCTGCCCTGTCTGG	4926
Db	2194	CGAACGGATGCGCAGAAAAAGTGTGGCGGGATCAGCGCGCTTGACGAGTCTGTTGG	2135
QY	4927	TATTACAGAGCTGCGATCTTCGGTGTGAACCTTCGCGTCTCGGCTGGCGGTCTACATGG	4986
Db	2134	TATCACGAACACGCGTATATGGGTCACCTTCGCGCGTAAGTACCCCTTTGTTATCGC	2075
QY	4987	TATCGGTACCGCAGCTATCGGTGGCGGTTTGTATTGTCACCTTTTGATATCAAGGAGTTCG	5046
Db	2074	CTGTATCAGTGGGCGCTTTGGGCGCACCATTAATGGCTACGCGCAACGAAAGTCTACTC	2015
QY	5047	GTTGGCGCTGCAAGTTCTTGGGTGTGTTTCTATTGATGCTCCAGATATGTCATGTT	5106
Db	2014	CTTTGGTTTGCAAGTATTTTCACTTCATGCAACCATCCCGCTCAAGGGAATGATTT	1955
QY	5107	CTTGGTTTGGCGGTAGTTACCTTTGTCAATCGCATTCGGCGCAGCGATTCCTTATGGCCT	5166
Db	1954	CACGCTGTGGGCGAGCGTTA---TTGGCGGTGTCAATGGCATCGGTTGCGCATTTGTCGG	1898
QY	5167	TTACTTGGTTGCGCGCAACGCGACATGATCCAGATGCAACCGCTGCTCCAGTGCCTGC	5226
Db	1897	TACGGTATGCTTCATTTTCATCCGCTAAACGCTACGCCAGC-----	1856
QY	5227	AGNAACACCAAGCCGAACGAGAACCCCGCAGAAATTTTCAACGATTCACCATCAT	5286
Db	1855	-GCAGGTTGCCCGAAGAGAAAAACACAGAGTTATTACACCCTGAGCAGGCGCGTAT	1797
QY	5287	CCAGGCACCTTTTACCCTGTAAGCATATCGCACTGAGCAGCGTCAAGCATGCCATGTTTCG	5346
Db	1796	CTGTTACCGATGACGGGAGAGATTGTGCTGCTATTCACGTCTGATACCAAGTTCG	1737
QY	5347	CAGCGGAAGCTTGGCTCAGTGTTCGATCTGCTCCCAACGAGGGCAGCTGTTTTCACC	5406
Db	1736	CAGTGGCCTGTTGGGTAAGGATATTGGCAATCTGCGCCTCGGTTGGTGAAGTGGTCTCC	1677
QY	5407	AGTGAGCGGAAGATCGTGTGGCCTTCCCATCTGTGTACGCTTTGCGATCTGCGCGCACTAA	5466

Db	1676	GGTGGGGTGGGAATTCCTCGTTGTCGCCAATACACGCCATTG-----GCATPTGA	1623
Qy	5467	GGCTGAGGATGGTTCCAAATCTGGATATCTTTGATGTCACCAATGGTTTCGACACCGTAAACCT	5526
Db	1622	GTCCAGATGATGG-----TGTGGAGATCCTGATTATGTCCGGTATCGACACCGTAAAACT	1569
Qy	5527	CAACGGCACGCATTTAAACCGCTGAGAAGACGCGGCGATGAAGTCAAAAGCAGGGGAGCT	5586
Db	1568	GGACGGGCAAAATCTTTTCCGCTCAGCTCAACGTTGGGTGACAAGTCAATACAGCGATCG	1509
Qy	5587	GCTGTGTGAATTCGATATTGATGCCATTAAAGGCTGCAGGTTATGAGGTAACACGCGGAT	5646
Db	1508	GCTGATTCTTTTGATATCCCTGCTATTTTCGAGGCGCGATTTGATCTGACGACGCGGT	1449
Qy	5647	TGTTGTTCCAAATACAGAAACCGGACCTGTAAACACTTACGGTTTGGCGGAAATGA	5706
Db	1448	ATTAATCAGTAATAGCGATGATTTTACGGACGATATACCCACGCGACGGCAGATAAAG	1389
Qy	5707	AGCGGGAGCCCAACCTGCTCAACGCT	5731
Db	1388	CGCAGGTGAACCGCTGTTATCCATC	1364

RESULT 11
 US-08-673-190A-3/c
 ; Sequence 3, Application US/08673190A
 ; Patent No. 5985668
 ; GENERAL INFORMATION:
 ; APPLICANT: Mattes, Ralf
 ; APPLICANT: Klein, Kathrin
 ; APPLICANT: Stegmair, Sabine
 ; TITLE OF INVENTION: Sucrose Metabolism Mutants.
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Finnegan, Henderson, Parabow, Garrett & Dunner
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/673,190A
 ; FILING DATE: 27-JUN-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Forman, David S.
 ; REGISTRATION NUMBER: 33,694
 ; REFERENCE/DOCKET NUMBER: 06473.0001-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)408-4000
 ; TELEFAX: (202)408-4400
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 465 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: Protaminobacter rubrum
 ; US-08-673-190A-3

Query Match 1.6%; Score 95; DB 2; Length 465;
 Best Local Similarity 53.3%; Pred. No. 1.2e-19;
 Matches 200; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qv 4414 CGCGGGCGAAATGCCATGTGGTCCCTGTTGGTGGATGTTGCTCAAGCTGGTTACCA 4473


```
Db 1312 GCTACAGGAATTCCTTTGTTCTGTCATTCGTAGTACATTTAGGATTTGACGAA 1371
Qy 5192 ATTGATCCAGATCAACCGCTGCTCCAGTCCCTGCGAGGACGACCAAGCGAAGCAGAA 5251
Db 1372 CAACACAGACAAAT-----CAATTAGAAACAACAATCGGAATGACGA 1416
Qy 5252 GCACCCGAGAAATTTCAACAGATTCACCATCATCCAGGACCTTTGACCGGTGAAGCT 5311
Db 1417 GAACCGATTACTTCAGACGCGCAT-----ACCTTAAAGTCTCTTAACGGTAAAGTA 1470
Qy 5312 ATGCGACTGAGCGCTCAGCGATGCCATGTTTCCAGCGGAAGCTTGGCTCAGGTGT 5371
Db 1471 CTGCCATTGAGTAAGTACACGACCAAGATTTTTCTTCTGGCTGATGGAAAGATATC 1530
Qy 5372 GCGATGTCGCCACCAAGGCGAGCTGGTTTCCAGTACGCGGAAGATCGTGGCGC 5431
Db 1531 GCTATTGATCCAGAAAGTAGGAGAACTAGTTGCACCTGCAGACGAGAGATCAGACTATT 1590
Qy 5432 TTCCCATCTGTCACGCTTTTCGAGTCCGACCTAAGCTGAGGATGTTCCATGTGAT 5491
Db 1591 TTCCCACTGGACATGCTGTAGGATCACACGA-----CTGACGGAGCGGAG 1638
Qy 5492 ATCTTGATGCATTTGGTTTCGACACCGTAAACCTCAACGGCAGCGACCTTTAACCGCTG 5551
Db 1639 ATTTGATTCATATCGGGATGGATACAGTAGAATTAATGGGAATGGCTTTGAGATCTTA 1698
Qy 5552 AAGAAGCGGCGATGAGTCAACGAGGAGCTGCTGTGTGAATTCGATATTGATGCC 5611
Db 1699 GTCAACAGGGGATCTCGTCAAGCAGGAGATTTATGATTCGTTTGCATCGAAGCA 1758
Qy 5612 ATTAAGCTCAGGTTATGAGTAAACCGCGGATTTGTTTCGAAT 5659
Db 1759 ATTAGACAGCTGGCTATAGCGTAATCACTCCAGTTGTCAATACGAAT 1806
```

RESULT 13

US-09-221-017B-736/c
; Sequence 736, Application US/09221017B
; Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 736:
SEQUENCE CHARACTERISTICS:
LENGTH: 2295 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...2295
US-09-221-017B-736

Query Match 1.5%; Score 92; DB 4; Length 2295;

Best Local Similarity 51.6%; Pred. No. 4.1e-18;

Matches 285; Conservative 0; Mismatches 260; Indels 7; Gaps 3;

```
Qy 1636 TGGGCTTGCACACGATCCCTCACCACGTAGTACCTACCAAGAGCTCATTCGCGATG 1695
Db 1642 TGGGCTTGCACACGATCCCTCACCACGTAGTACCTACCAAGAGCTCATTCGCGATG 1583
Qy 1696 AAGCTGGGAAAGTGTCTTCAAGAACTGCAAGGATCTTCTTGGATGAATACGTGGGAC 1755
Db 1582 AAGAAGGATGCATATCTTTCCGCCACGTAATCACCTTCAACATGATGATACGTAGGCC 1523
Qy 1756 TAACCCGTGACGATGAACACAGCTACTTTAAACCATTCGCAAGAGTTCACCTACCA 1815
Db 1522 TTGCTATAGAGCATCCGAGAGCTATCATTCATGTCATCGGCATTTGTTCCGACCAT 1463
Qy 1816 TCGACATCGTTGATGAAGAGGTCTACAGCCAGATGTTGCAAAACCCCTGATC---CATACG 1872
Db 1462 TCGACATCGTCTCCTCAGAACATCCACATTCCTCAATGGCAATGCTCTGACCTGACGCCG 1403
Qy 1873 AAGCAGCTGCAGAGTATGAGCAAGATCGCTGCAGATCCGTTGAGTTCAAATCCTTG 1932
Db 1402 AGTGCATGCATACGAGAGAGCTATTGAAGCTGCCGAGGATAGACCTTTTCATCGGCG 1343
Qy 1933 GCATCGCGGAAACGG--CACATCGCTTTTCATTGAACCATCATCTTCTCTGTCAGGACTGA 1991
Db 1342 GGATAGGACCTCAGCGACATATCGCATTCATGAGCCGGGTCTCTCTTACCTCTCGCA 1283
Qy 1992 CAAAGTCCAGCGCTGCACCTTAAACTGTGGAGGACACGCTCGATTCTTCAACACC- 2050
Db 1282 CCCGTATCAAGACGCTAACTACAGATACCCGTACTGGCTAAACAGTCGCTTTTTCGACCA 1223
Qy 2051 --ATCGAAGAGGTCCCAACCCACGCGCTACCCAGGGTTTGGGACACTTTGTCGCCGCGC 2108
Db 1222 ATACGAATCAGGTTCCGAAACGTGCTCTCACCGTAGGATAGGAACATATTGATGCAC 1163
Qy 2109 AAAACATCGTTGGTGGCACTGTGTGAAGAAAGCGGACGCCATCCGCGGAAGTGTGG 2168
Db 1162 GGAAGTAATGATCTCTCGTGAACGACATACTAAAGCCCGTCTCTTCTGTAAGCTGTGG 1103
Qy 2169 AAGGCCAGTGA 2180
Db 1102 AAGTGCAGTCA 1091
```

RESULT 14

US-09-367-293-1

; Sequence 1, Application US/09367293

; Patent No. 6444878

; GENERAL INFORMATION:

APPLICANT: Donaldson, Lain A.
APPLICANT: Bolsen, Kirsten
APPLICANT: Jorgensen, Kirsten
APPLICANT: Jorsboe, Morten
TITLE OF INVENTION: SELECTION METHOD FOR TRANSGENIC PLANTS
FILE REFERENCE: DYO21.00LAPC
CURRENT APPLICATION NUMBER: US/09/367,293
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: PCT/GB98/00367
PRIOR FILING DATE: 1998-02-05
PRIOR APPLICATION NUMBER: GB 9702592.8
PRIOR FILING DATE: 1997-02-07
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 834
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (0)...(0)
OTHER INFORMATION: Glucosamine-6-Phosphate Deaminase
US-09-367-293-1
Query Match 1.5%; Score 90.6; DB 4; Length 834;
Best Local Similarity 48.7%; Pred. No. 5e-18;
Matches 309; Conservative 0; Mismatches 319; Indels 7; Gaps 2;
QY 1636 TGGGGCTTGCACAGGATCTCCACCTAGTACCTACCAAGAGCTCATTCGCGATGTATG 1695
DB 125 TGGGCTTGCACAGGATCTCCACCTAGTACCTACCAAGAGCTCATTCGCGATGTATG 184
QY 1696 AAGCTGGGAGTGTCTTCAAGACCTGCAAGGATCTTCTTGGGATGATAGTGGGAC 1755
DB 185 AAGCAGGCGGAGTGTCTTCAAGACCTGCAAGGATCTTCTTGGGATGATAGTGGGAC 244
QY 1756 TAACCGGTGAGTGTCTTCAAGACCTGCAAGGATCTTCTTGGGATGATAGTGGGAC 1815
DB 245 TGGGCTTGCACAGGATCTCCACCTAGTACCTACCAAGAGCTCATTCGCGATGTATG 304
QY 1816 TCGACATCGTGTGATGAGAGGTCTACAGCCAGATGTGTGCAACCTGTGATCCATACGAG 1875
DB 305 TTGATATTCACGAGAGAACTCAACCTTCTCAACGCGACCGCCCGGATATCGAGCGG 364
QY 1876 CAGCTGCAGAGTATGAGGAGGATCGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1935
DB 365 AGTCCGCGGAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 424
QY 1936 TCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1991
DB 425 GTGTAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 484
QY 1992 CAAAGGTCCAGGCGGTGACCCCTAAAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2048
DB 485 CTCGTATCAAAACCTGACTCATGACACTCGCTGCGCAAACTCTCGTTCTTTGATACG 544
QY 2049 CCATGGAAGAGGTCCCAACCCGCGTCAACCCAGGTTTGGGAGGAGGAGGAGGAGGAGGAGG 2108
DB 545 ATGTATATCAGGTGCAAAATATGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 604
QY 2109 AARACATCGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2168
DB 605 AAGAGTGTGATGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 664
QY 2169 AAGGCCAGTGTGATCTTCTTGGCCAGGTTCCATCTGTAGATGCAACATGCGACCAT 2228
DB 665 AAGGTGCGGTGAACCATATGTGGACCATGCTGTGTGCAACTGCTGTGCAACTGCTGTGCA 724
QY 2229 CATCGTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2263
DB 725 TGGTGTGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 759

RESULT 15
US-09-557-884-1/c
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1
Query Match 1.5%; Score 90; DB 4; Length 1830121;
Best Local Similarity 49.5%; Pred. No. 4.3e-15;
Matches 291; Conservative 0; Mismatches 290; Indels 7; Gaps 2;
QY 1635 TTGGGCTTGCACAGGATCTCCACCTAGTACCTACCAAGAGCTCATTCGCGATGTAT 1694
DB 156836 TTGGGCTTGCACAGGATCTCCACCTAGTACCTACCAAGAGCTCATTCGCGATGTAT 156777
QY 1695 GAAGCTGGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 1754
DB 156776 CAAGCAGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 156717
QY 1755 CTAAACCGGTGACGATGAAACAGAGTACTTTAAACACCATTCGCAAGAGTGTCTACTGACAC 1814
DB 156716 TTACCAGAGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 156657
QY 1815 ATCGACATCGTGTGATGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 1874
DB 156656 ATCGATATTTTACCTGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 156597
QY 1875 GCAGCTGCGAGAGTGTGAGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 1930
DB 156596 GAATGCGGCTGTGTATGAGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 156537
QY 1931 TGGCATCGGCGGAGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 1990
DB 156536 GGGGTTGGGCTGTGTATGAGGAGGAGTGTCTTCAAGAACTGCAAGGAGTGTCTTGTGATGAATACGTGGCA 156477

Wed Oct 1 09:08:28 2003

```
QY 1991 ACAAGGTCAGGCGCTGCACCTAAACTGTGGAGACAAAGCTCGATTCTTCAACA-- 2048
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156476 ACCCGTATTAAACGCTAACGCAAGATACATTGATTGCAAAATTCACGCTTCITTTAATAAT 156417
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2049 -CCATCGAAGAGGTCCCAACCCACCGCGTCACCCAGGGTTTGGGCACCTTGTCCCGCGG 2107
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156416 GATGTTACTCAAGTGCCGAATATGCTTTAACCATTTGGTGTGGCACATTATTAGATGCA 156357
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2108 CAAAACATCGTGTGTGGCAACTGGTGAAGGAAAGCCGACGCCATCCGCGGAACGTGTG 2167
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156356 GAAGAAGTGATGATTTAGCAACAGGACATCAAAAAGCACTGGCAGTCAAGCTGCAGTT 156297
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2168 GAAGGCCCAAGTACTGCTTCTTGGCCAGGTTCCATCCTGTAGATGCAC 2215
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156296 GAAGGCAGTATTAAACCATTTATGGACAGTCAGTGCATTCACAAATGCAC 156249
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: October 1, 2003, 01:26:45

Job time : 282 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 30, 2003, 23:16:34 ; Search time 939 Seconds
(without alignments)
15825.103 Million cell updates/sec

Title: US-10-019-284A-1
Perfect score: 5969
Sequence: 1 agccctgcagccaccatt.....tgatatattgcccgcgtgaa 5969

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1078620 seqs, 1244745471 residues

Total number of hits satisfying chosen parameters: 3357240

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	5656.2	94.8	3309400	10	US-09-738-626-1 Sequence 1, Appli
2	1874.2	31.4	1983	10	US-09-738-626-2904 Sequence 2904, Ap
3	1267.8	21.2	1299	10	US-09-738-626-2905 Sequence 2905, Ap
4	1110.4	18.6	1152	10	US-09-738-626-2907 Sequence 2907, Ap
5	671.8	11.3	759	10	US-09-738-626-2906 Sequence 2906, Ap
C 6	215.2	3.6	1059	11	US-09-746-660A-31 Sequence 31, Appl
7	192	3.2	975	10	US-09-974-300-829 Sequence 829, App
8	177.6	3.0	30246	8	US-08-781-986A-56 Sequence 56, Appl
9	174.2	2.9	1380	10	US-09-974-300-685 Sequence 685, App
10	163.6	2.7	1884	9	US-09-815-242-9369 Sequence 9369, Ap
11	161	2.7	1380	10	US-09-738-626-2903 Sequence 2903, Ap
12	153	2.6	5840	10	US-09-070-927A-31 Sequence 31, Appl
13	122.2	2.0	567	10	US-09-070-927A-664 Sequence 664, App
14	121.2	2.0	1143	14	US-09-070-927A-31 Sequence 31, Appl
15	121.2	2.0	9025608	14	US-10-156-761-1 Sequence 1, Appli
C 16	118.4	2.0	936	10	US-09-738-626-2908 Sequence 2908, Ap

17	117.4	2.0	783	14	US-10-156-761-3005 Sequence 3005, Ap
18	116	1.9	1098	10	US-09-974-300-734 Sequence 734, App
19	109.8	1.8	1014	10	US-09-974-300-822 Sequence 822, App
20	98.2	1.6	24004	10	US-09-070-927A-72 Sequence 72, Appl
21	97.2	1.6	882	10	US-09-974-300-5255 Sequence 5255, Ap
22	96.4	1.6	489	10	US-09-974-300-857 Sequence 857, App
C 23	90	1.5	1830121	14	US-10-329-960-1 Sequence 1, Appli
24	88.8	1.5	722	10	US-09-974-300-814 Sequence 814, App
C 25	86.4	1.4	7156	10	US-09-070-927A-29 Sequence 29, Appl
26	82.4	1.4	348	10	US-09-974-300-5313 Sequence 5313, Ap
27	79	1.3	480	11	US-09-918-995-21116 Sequence 21116, A
28	77.4	1.3	444	10	US-09-974-300-877 Sequence 877, App
29	77.2	1.3	474	10	US-09-070-927A-639 Sequence 639, App
30	77	1.3	1353	10	US-09-974-300-799 Sequence 799, App
31	76.8	1.3	1368	10	US-09-974-300-893 Sequence 893, App
32	75.6	1.3	1347	9	US-09-741-669-137 Sequence 137, App
33	75.6	1.3	1947	9	US-09-815-242-5992 Sequence 592, Appli
C 34	74.2	1.2	9025608	14	US-10-156-761-1 Sequence 1, Appli
35	73.2	1.2	465	8	US-08-781-986A-1317 Sequence 3951, Ap
36	73	1.2	1398	14	US-10-156-761-3951 Sequence 6, Appli
37	72.2	1.2	2130	9	US-09-079-892-6 Sequence 647, App
C 38	71.8	1.2	1216	14	US-10-106-698-647 Sequence 802, App
39	71.6	1.2	722	10	US-09-974-300-892 Sequence 550, App
C 40	71.2	1.2	9797	10	US-09-070-927A-550 Sequence 1956, Ap
41	69.6	1.1	390	8	US-08-781-986A-1958 Sequence 5262, Ap
C 42	68.2	1.1	405	10	US-09-974-300-5262 Sequence 5182, Ap
43	67.8	1.1	429	10	US-09-974-300-5182 Sequence 849, App
44	67.6	1.1	612	10	US-09-974-300-849 Sequence 153, App
45	64.4	1.1	1314	9	US-09-939-980-153 Sequence 153, App

ALIGNMENTS

RESULT 1

US-09-738-626-1/C
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 94.8%; Score 5656.2; DB 10; Length 3309400;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 5793; Conservative 0; Mismatches 173; Indels 3; Gaps 3;

QY 1 AGTCGTCGACGCCACCATTTGCTGCTACCGAGCTTGGAGGCTTTTACATCT 60
|||||

Db	2815585	AGTCGCTCGAGCCACCATTGATGAGGTGGTCAACAGAGCTTGGGAGGCTTTCTACATCC	2815526
Qy	61	ACGCTCCCGTCGGCGTGGAGTGGGTCATTACGGGTGGATCACGCCGTGAAAGTGGC	120
Db	2815525	ACACTGCCGTGGCGTGGAGTGGGTCATTACGGGTGGATCACGCCGTGAAAGTGGC	2815466
Qy	121	GAACCCATGATGTTCCCTTGGGTTGAGGAACGAGTGGGTTGAAAGTTTTTCAGNG	180
Db	2815465	GAACCCATGATGTTCCCTTGGGTTGAGGAACGAGTGGGTTGAAAGTTTTTCAGNG	2815406
Qy	181	TCGCAATTTTAAAGTTATGCATCATCAGCTTGGAAAGGTGAGTTAAATCAGTAGACCTG	240
Db	2815405	TCGCAATTTTAAAGTTATGCATCATCAGCTTGGAAAGGTGAGTTAAATCAGTAGACCTG	2815346
Qy	241	CACACAGCCCTCAAGTCGGAAGATAATTAACCTAGATCCGAGACATAAGACATCA	300
Db	2815345	CACACAGCCCTCAAGTCGGAAGATAATTAACCTAGATCCGAGACATAAGACATCA	2815286
Qy	301	CGTCTATGCTTCTGCTGAAGAACCAATAAATCAGTCAAGAAAGTGGCAGAGTGGTGCATT	360
Db	2815285	CGTCTATGCTTCTGCTGAAGAACCAATAAATCAGTCAAGAAAGTGGCAGAGTGGTGCATT	2815226
Qy	361	ATCAAGAAATGAGGTCACAGCTTAAAGAAATGAGGAAAGTGGTCCCGCTCG	420
Db	2815225	ATCAAGAAATGAGGTCACAGCTTAAAGAAATGAGGAAAGTGGTCCCGCTCG	2815166
Qy	421	GGGTGATGATGCTTCTCCAACTCGAAACGCGATCATCACGGAATCTCTGGAGA	480
Db	2815165	GGGTGATGATGCTTCTCCAACTCGAAACGCGATCATCACGGAATCTCTGGAGA	2815106
Qy	481	CAGCACTTAAAGCGAGGATTCACCCCGAATCCACAGATGTTCCCGTTTATG	540
Db	2815105	CAGCACTTAAAGCGAGGATTCACCCCGAATCCACAGATGTTCCCGTTTATG	2815046
Qy	541	ATCTTCAATACAGGTGGAACGGTGGCGCTTCTACGGAAACGACGACGCGA	600
Db	2815045	ATCTTCAATACAGGTGGAACGGTGGCGCTTCTACGGAAACGACGACGCGA	2814986
Qy	601	GGACACCGCGAGTATCACGGGAACATGGCAGCAGCTGATGTTCCCGAAGCATGTTT	660
Db	2814985	GGATGCGCGGAGTATCACGGGAACATGGCAGCAGCTGATGTTCCCGAAGCATGTTT	2814926
Qy	661	CGGCGCGGCTGACGCACTGCGCAGCGAGGTGGAACCTTATCCCTTGTGGAAGG	720
Db	2814925	CGGCGCGGCTGACGCACTGCGCAGCGAGGTGGAACCTTATCCCTTGTGGAAGG	2814866
Qy	721	TCCTGCTGCGGCAATTCACCTCGAGGCGCTTTCATCAACGATGCGGTTGTTGCTC	780
Db	2814865	GCCTGCTGCGGCAATTCACCTCGAGGCGCTTTCATCAACGATGCGGTTGTTGCTC	2814806
Qy	781	AAACCCGGATTCATTTTCCCGCAACCCACAGATCTTCCCGGCTGATCCATCGG	840
Db	2814805	AAACCCGGATTCATTTTCCCGCAACCCACAGATCTTCCCGGCTGATCCATCGG	2814746
Qy	841	GAAAGGTTGGATCAATTCGATCAGATAGCGCGGAACTGACAATCTTCTGAGCTC	900
Db	2814745	GAAAGGTTGGATCAATTCGATCAGATAGCGCGGAACTGACAATCTTCTGAGCTC	2814686
Qy	901	TCGATCTCTGCGAGCGCACATCATCTGTTCCCTTCGCGCACACTGATGCAATTTG	960
Db	2814685	TCGATCTCTGCGAGCGCACATCATCTGTTCCCTTCGCGCACACTGATGCAATTTG	2814626
Qy	961	ATACCACTTACCGCGCAATTCGCTTAAAGAGAAAAATGTGACGCTCACGCTACGC	1020
Db	2814625	ATACCACTTACCGCGCAATTCGCTTAAAGAGAAAAATGTGACGCTCACGCTACGC	2814566
Qy	1021	ATTTGTTCAATGCGATGCTCCGCTGCATCATAGGGCTCCGCGAGCGTGGCGCTTGC	1080
Db	2814565	ATTTGTTCAATGCGATGCTCCGCTGCATCATAGGGCTCCGCGAGCGTGGCGCTTGC	2814506
Qy	1081	TGCTGCGGACGCTGCGGCGACATATGTTGAGTTGATCCGCGAGCGGCTGCATTGG	1140
Db	2814505	TGCTGCGGACGCTGCGGCGACATATGTTGAGTTGATCCGCGAGCGGCTGCATTGG	2814446
Qy	1141	CCGATGGAACGGTGCATCTAGCTCGTTCCCAACAGCCTTTTTCATCACGAGCCCATGG	1200
Db	2814445	CCGATGGAACGGTGCATCTAGCTCGTTCCCAACAGCCTTTTTCATCACGAGCCCATGG	2814386
Qy	1201	AAGCCGCGGAATGCCACACGGTGAATACATTTTGGGCGTTTGAACGTCACCGTCACG	1260
Db	2814385	AAGCCGCGGAATGCCACACGGTGAATACATTTTGGGCGTTTGAACGTCACCGTCACG	2814326
Qy	1261	ATGGAGTCCGCGCTCGGCGATGGCGGCATCGCGGGGACACGACACATAGCGA	1320
Db	2814325	ATGGAGTCCGCGCTCGGCGATGGCGGCATCGCGGGGACACGACACATAGCGA	2814266
Qy	1321	GTCACTTCGTGCACACGCTGGCGAGGGTATGAGCTTATCGAGCGGACCTCCACACT	1380
Db	2814265	GTCACTTCGTGCACACGCTGGCGAGGGTATGAGCTTATCGAGCGGACCTCCACACT	2814206
Qy	1381	CAACCGTCCGCGCTTAAATTTCTCGTCTTGGGATCAGAAATCGTAAATCCACCTG	1440
Db	2814205	CAACCGTCCGCGCTTAAATTTCTCGGACTTAGCATCAGAAATCGTAAATCCACCTG	2814146
Qy	1441	CAATTTTGTGCTTTTGTACTCAACGGCCAGGTGCAAAAGTCCATTTAGTGCATCAAG	1500
Db	2814145	CAATTTTGTGCTTTTGTACTCAACGGCCAGGTGCAAAAGTCCATTTAGTGCATCAAG	2814086
Qy	1501	TACTTTAAGTAGAGTAAACTATCTCTGAT-TTTAAAGAGTCCACCATGGAATCACT	1559
Db	2814085	TACTTTAAGTAGAGTAAACTATCTCTGAT-TTTAAAGAGTCCACCATGGAATCACT	2814026
Qy	1560	ATCTGCAAGACGACGAGAGTTCGCAAGCAGTTGCAAGTCTTAAATCGACCTTCGCC	1619
Db	2814025	ATCTGCAAGACGACGAGAGTTCGCAAGCAGTTGCAAGTCTTAAATCGACCTTCGCC	2813966
Qy	1620	AACAGGGTGAACCTTGGGGTTCACACAGAGTCTCACACTGAGTACCTACCAAG	1679
Db	2813965	ACTAAGGGCGGAACCTTGGGGTTCACACTGAGTCTCACACTGAGTACCAAG	2813906
Qy	1680	CTCATTCGATGATGAACTGGGAAGTGTCTTCAAGAACTGCAAGGATCTTGTGTTG	1739
Db	2813905	CTCATTCGATGATGAACTGGGAAGTGTCTTCAAGAACTGCAAGGATCTTGTGTTG	2813846
Qy	1740	GATGAATACGTGGGACTAACCCGTGACGATGAAACACAGTCTTTTAAACCATTCGCAA	1799
Db	2813845	GATGAATACGTGGGACTAACCCGTGACGATGAAACACAGTCTTTTAAACCATTCGCAA	2813786
Qy	1800	GAGTTCACTGACCATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTGCAAC	1859
Db	2813785	GAGTTCACTGACCATCGACATCGTTGATGAAGAGTCTACAGCCAGATGGTGCAAC	2813726
Qy	1860	CCTGATCCATACGACGCTGACGATGATGAGGCAAGATCGTGCAGATCCGTTGAA	1919
Db	2813725	CCTGATCCATACGACGCTGACGATGATGAGGCAAGATCGTGCAGATCCGTTGAA	2813666
Qy	1920	GTTCAATCCTTGGCATCGCGGAAACGG-CACATCGCTTTCATTTGAACCATCTCTCT	1978
Db	2813665	GTTCAATCCTTGGCATCGCGGAAACGG-CACATCGCTTTCATTTGAACCATCTCTCT	2813606
Qy	1979	CTGTCAGGACTGACAAAGTCCAGCGCTGCAACCTTAAACCTGTGAGGACAACTCGA	2038
Db	2813605	CTGTCAGGACTGACAAAGTCCAGCGCTGCAACCTTAAACCTGTGAGGACAACTCGA	2813546
Qy	2039	TTCCTTCAACACCATCGAAGAGTCCCAACCCAGCGGTCCACCGAGTTTGGGACTTGG	2098
Db	2813545	TTCCTTCAACACCATCGAAGAGTCCCAACCCAGCGGTCCACCGAGTTTGGGACTTGG	2813486
Qy	2099	TCCCGCGCAAAACATCTGTTGTTGGCAACTGTTGAAGGAAAGCCGACCCATCCGC	2158
Db	2813485	TCCCGCGCAAAACATCTGTTGTTGGCAACTGTTGAAGGAAAGCCGACCCATCCGC	2813426
Qy	2159	GGAACGTGTGAAGGCGGACATGACTGCTTCCGACAGTTTCCATCTGTAGATGACAC	2218
Db	2813425	GGAACGTGTGAAGGCGGACATGACTGCTTCCGACAGTTTCCATCTGTAGATGACAC	2813366

QY 2219 ATGCCACCATCATCGTTGGATGAAGCAGCAGATCCAAAGCTTGGAAACGCTGATCACTAC 2278
Db 2813365 AATGCCACCATCATCGTTGGATGAAGCAGCAGATCCAAAGCTTGGAAACGCTGATCACTAC 2813306
QY 2279 CGTCTCATGAGCAATTAAGCTGCGCTAGAAACAAAAGAAAGTACTGTGTGGGGCTA 2338
Db 2813305 CGTCTCATGAGCAATTAAGCTGCGCTAGAAACAAAAGAAAGTACTGTGTGGGGCTA 2813246
QY 2339 TGCACACAGAACTTCCAGTTTGGCCCTCGCTGCTGACATGTGACTCTCTCCGACGGGACGC 2398
Db 2813245 TGCACACAGAACTTCCAGTTTGGCCCTCGCTGCTGACATGTGACTCTCTCCGACGGGACGC 2813186
QY 2399 TCAATGATCCCAACGGGAATGACTGCTGATGGAGATACCCCTCCACGCTCTACTACACGACG 2458
Db 2813185 TCAATGATCCCAACGGGAATGACTGCTGATGGAGATACCCCTCCACGCTCTACTACACGACG 2813126
QY 2459 ATCCAGGTTTCCCTTCCGACCAAAAGCGACCGCTGGGCTCACACCAACGACCGGTTGA 2518
Db 2813125 ATCCAGGTTTCCCTTCCGACCAAAAGCGACCGCTGGGCTCACACCAACGACCGGTTGA 2813066
QY 2519 CCGGACCGCAGCGATTGCAGTGGACGACGACCTGCCGACGCTCTTTACCCGGATGCATCCT 2578
Db 2813065 CCGGACCGCAGCGATTGCAGTGGACGACGACCTGCCGACGCTCTTTACCCGGATGCATCCT 2813006
QY 2579 ATGACCTGGATGGATGCTATTCGGGTGGAGCCGCTATTACTGACGGCACACTTAACTTT 2638
Db 2813005 ATGACCTGGATGGATGCTATTCGGGTGGAGCCGCTATTACTGACGGCACACTTAACTTT 2812946
QY 2639 TCTACACGGCAACTTAAATTCACGGAAGCGCGCCGACCCAAAACCTTGTCCAAG 2698
Db 2812945 TCTACACGGCAACTTAAATTCACGGAAGCGCGCCGACCCAAAACCTTGTCCAAG 2812886
QY 2699 TCGAGGACCCAACTGGGCTGATGGGGCGCATTCATCGCGCTTCGCTTAAATTCGCTTA 2758
Db 2812885 TCGAGGACCCAACTGGGCTGATGGGGCGCATTCATCGCGCTTCGCTTAAATTCGCTTA 2812826
QY 2759 TCGAGGACCCGCGCAGGGTTTACACCCCATACCGCGATCCCATGATCAGCCCTGATG 2818
Db 2812825 TCGAGGACCCGCGCAGGGTTTACACCCCATACCGCGATCCCATGATCAGCCCTGATG 2812766
QY 2819 GTGATGGTTGAAACATGGTTCTTGGGGCCCAACGCGAAACCTCACCGGTGACGGGTTT 2878
Db 2812765 GTGATGGTTGAAACATGGTTCTTGGGGCCCAACGCGAAACCTCACCGGTGACGGGTTT 2812706
QY 2879 TATACCGCTCCACAGATCTTGAAACCTGGGAATTCCTCGGTTGAAATCACCTTTCACCTCA 2938
Db 2812705 TATACCGCTCCACAGATCTTGAAACCTGGGAATTCCTCGGTTGAAATCACCTTTCACCTCA 2812646
QY 2939 GTGATGCAACACCTGGTTCTGCTCTGATCTGCTGATCTGCTGATGCTGATGCTGATGCT 2998
Db 2812645 GTGATGCAACACCTGGTTCTGCTCTGATCTGCTGATCTGCTGATGCTGATGCTGATGCT 2812586
QY 2999 CCAACCTTTTACGCTTCGCGATGAAGAACTGGCGAGATCTCGAGTGTGCTGATTTCT 3058
Db 2812585 CCAACCTTTTACGCTTCGCGATGAAGAACTGGCGAGATCTCGAGTGTGCTGATTTCT 2812526
QY 3059 GTCCACAAAGGATTGGACCGAATCCACGATGAGGTTTACTCACTACGCAAGCTCTGACCACT 3118
Db 2812525 GTCCACAAAGGATTGGACCGAATCCACGATGAGGTTTACTCACTACGCAAGCTCTGACCACT 2812466
QY 3119 GCGGATATGTGCTCCACAGCTTGAAGAAAGCACTTCCGCGTCTTGGAGGATTCACCG 3178
Db 2812465 GCGGATATGTGCTCCGCAAGCTTGAAGAAAGCACTTCCGCGTCTTGGAGGATTCACCG 2812406
QY 3179 AGCTGGATTTCCGCGCATGAATTTACGCAACCGCAGGTTGCASTAAACGGTTTCTGATGCT 3238
Db 2812405 AGCTGGATTTCCGCGCATGAATTTACGCAACCGCAGGTTGCASTAAACGGTTTCTGATGCT 2812346
QY 3239 GCGCTGGGCTGGATGGGGCTGCGCCGCGAGGATGATCACCCAAACATTTGCAACAGGAAG 3298
Db 2812345 GCGCTGGGCTGGATGGGGCTGCGCCGCGAGGATGATCACCCAAACATTTGCAACAGGAAG 2812286
QY 3299 GATGGGTGCACTGCCTGACTGTGTCGCCGCCCAAGCTTCAITTTGGCAACACCGCATCTACC 3358

Db 2812285 GATGGGTGCACTGCCTGACTGTGCCCGCAAGCTTCAITTTGCGCAACACCGATCTATC 2812226
QY 3359 AAGAGCTCTTCTCCAGAGGGGAGTCGGGGTAAATCAGATCTGTATAGTTCTGAAC 3418
Db 2812225 AAGAGCTCTTCTCCAGAGGGGAGTCAGGGTAAATCAGATCTGTATAGTTCTGAAC 2812166
QY 3419 CTGTCCGAGTACACATCCGAGGCAATATTTCCCTCGAGTGGGATGCTGCTGTTTGTCTG 3478
Db 2812165 CTGTCCGAGTACACATCCGAGGCAATATTTCCCTCGAGTGGGATGCTGCTGTTTGTCTG 2812106
QY 3479 TGSATCGTATGGTGTATCGTCCGCTAGCTGAGGTAAAACCTGGCGCAATTAAGTATCGCGG 3538
Db 2812105 TGSATCGTGGTGTATCGTCCGCTAGCTGAGGTAAAACCTGGCGCAATTAAGTATCGCGG 2812046
QY 3539 ACGATAATACACGCTATTGAGATAAATGCAAGTGTGAGCAAGTTTCATTCGCTTTTCGCG 3598
Db 2812045 ACGATAATACGCTATTGAGATAAATGCAAGTGTGAGCAAGTTTCATTCGCTTTTCGCG 2811987
QY 3599 GCCTTCAAAGGTGACACTATTGAGAGTAAGTCAATATAAAGGGTCTTTTGTGGCGAATT 3658
Db 2811986 GCCTTCAAAGGTGACACTATTGAGAGTAAGTCAATATAAAGGGTCTTTTGTGGCGAATT 2811927
QY 3659 GTACAAATACTTCGCAAAATCCCTTGATCGGACACAAATAAAGGTTTAATATTGTTTA 3718
Db 2811926 GTACAAATACTTCGCAAAATCCCTTGATCGGACACAAATAAAGGTTTAATATTGTTTA 2811867
QY 3719 GCTTTTGAACAAACATTCATGTCTGAATTTTGTCTTCCCGGTGAAGAGAAATTC 3778
Db 2811866 GCTTTTGAACAAACATTCATGTCTGAATTTTGTCTTCCCGGTGAAGAGAAATTC 2811807
QY 3779 ATGGACCAATAAGGACCTCCGCAACGATCTCGCGGACATTTGGGGGCAACACACAT 3838
Db 2811806 ATGGACCAATAAGGACCTCCGCAACGATCTCGCGGACATTTGGGGGCAACACACAT 2811747
QY 3839 GTCCGCGCGCACACTGTCAACGCGTTTACGCTCTGCTCAAAGACACCAAGGATGTG 3898
Db 2811746 GTCCGCGCGCACACTGTCAACGCGTTTACGCTCTGCTCAAAGACACCAAGGATGTG 2811687
QY 3899 GATCGCAAAAGTGTGATGATCCAGATCTGAAAGGACCTTTGAAACTGGCGCATG 3958
Db 2811686 GATCGCAAAAGTGTGATGATCCAGATCTGAAAGGACCTTTGAAACTGGCGCATG 2811627
QY 3959 TTCCAGATCATCTCGGGCCAGCGATGTGGATCATGTTTTCAAAGAACTCGATGACGCA 4018
Db 2811626 TTCCAGATCATCTCGGGCCAGCGATGTGGATCATGTTTTCAAAGAACTCGATGACGCA 2811567
QY 4019 ACCTCCAAAGACATCGCTGTGTCCAGAGCAGCTCAAAGATGTTGCTGCTACACGCG 4078
Db 2811566 ACCTCCAAAGACATCGCTGTGTCCAGAGCAGCTCAAAGATGTTGCTGCTACACGCG 2811507
QY 4079 AACTGGTTACGCGGTGTGTGAAGGATTTGGCGGACATTTTCGCCGCTGATTCGAATC 4138
Db 2811506 AACTGGTTACGCGGTGTGTGAAGGATTTGGCGGACATTTTCGCCGCTGATTCGAATC 2811447
QY 4139 TTGGTTGGTGGCGGTGTGCTCATGCTATCAACAATGTGTGTTGGTGGCGAGGATCTGTC 4198
Db 2811446 TTGGTTGGTGGCGGTGTGCTCATGCTATCAACAATGTGTGTTGGTGGCGAGGATCTGTC 2811387
QY 4199 GGTCCGCAATCACTGTTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 4258
Db 2811386 GGTCCGCAATCACTGTTGGAGATGTTCCCTCAGATCAGCGGTGTTGCTGAGATGATCAAC 2811327
QY 4259 CTGATGCACTGCTGGCGGTTCGCGAGTTCCTCCAGTGTGTTGGTTTCCCGCAACCAAG 4318
Db 2811326 CTGATGCACTGCTGGCGGTTCGCGAGTTCCTCCAGTGTGTTGGTTTCCCGCAACCAAG 2811267
QY 4319 CGTTTCGGTGGCAATGAGTTCCCTGGGCGCGGATTTGGTATGCGGATGTTGTTCCCAAC 4378
Db 2811266 CGTTTCGGTGGCAATGAGTTCCCTGGGCGCGGATTTGGTATGCGGATGTTGTTCCCG 2811207
QY 4379 CTGGTTAACGGCTACGACGTGGCGCGCAACCATGACCGCGGCGAAATGCCAATGTGCTCC 4438

Db 2811206 TTGGTGAACGGGTACGACGTGGCCGCCACCATGGCTGCGGGGGAATGCCAATGTGGTCC 2811147
QY 4439 CTGTTGGTTTGGATGTGCTCAAGCTGTTTACAGGACCGTCTTCCTGTGCTGGT 4498
Db 2811146 CTGTTGGTTTGGATGTGCTCAAGCTGTTTACAGGACCGTCTTCCTGTGCTGGT 2811087
QY 4499 GTCTCTTGATCTGCAACGATGAGAGTTCTTGCACAGGACTCATGGGCACTGCA 4558
Db 2811086 GTTCTTGATCTGCAACGATGAGAGTTCTTGCACAGGACTCATGGGCACTGCA 2811027
QY 4559 GACTTCTGATCACCCAGTGTGTGACTGTGCTGCTCACCGGTTCTTACGTTTCACTT 4618
Db 2811026 GACTTCTGATCACCTCCAGTGTGCTGCTGCTCACCGGTTCTTACGTTTCACTT 2810967
QY 4619 ATTGTCCAGCATCGCTGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4678
Db 2810966 ATTGGCCAGCATCGCTGGTGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2810907
QY 4679 GATTTCGGTGGTCCAGTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4738
Db 2810906 GATTTCGGTGGTCCAGTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2810847
QY 4739 ACTGGTCTGACAGTCCCTTCCCGCCCAATGAGCTGGAGCTGTTTAAACAGGTTGGATCC 2810787
Db 2810846 ACTGGTCTGACAGTCCCTTCCCGCCCAATGAGCTGGAGCTGTTTAAACAGGTTGGATCC 2810787
QY 4799 TTCACTTTCGCAACCGCATCCATGCGCAATATCGCGCAGGTTGCAAGCTGTTTGGCAGTG 4858
Db 2810786 TTCACTTTCGCAACCGCATCCATGCGCAATATCGCGCAGGTTGCAAGCTGTTTGGCAGTG 2810727
QY 4859 TTTCTTCCAGCAAGAGTGAAGAGTCAAGGCTTTCAGAGCTGCTTTCAGGTTCTCCGCT 4918
Db 2810726 TTTCTTCCAGCAAGAGTGAAGAGTCAAGGCTTTCAGAGCTGCTTTCAGGTTCTCCGCT 2810667
QY 4919 GTTCTGGTATTACAGAGCTGCGATCTTCGGTGTGAACCTTCGGCTGCGCTGGCGCTTC 4978
Db 2810666 GTTCTGGTATTACAGAGCTGCGATCTTCGGTGTGAACCTTCGGCTGCGCTGGCGCTTC 2810607
QY 4979 TACATGTTATGCTGACGAGCTATCGGTGGGCTTTCGATGCTCTTTCGATATCAAG 5038
Db 2810606 TTCATCGGTATCGGTATCCGAGCTATCGGTGGGCTTTCGATGCTCTTTCGATATCAAG 2810547
QY 5039 GCAGTTGGCTTGGGCTGCTGAGTTCCTTGGGCTTTCGATGCTCTTTCGATATCAAG 5098
Db 2810546 GCAGTTGGCTTGGGCTGCTGAGTTCCTTGGGCTTTCGATGCTCTTTCGATATCAAG 2810487
QY 5099 GTCATGTTCTGTTGGGCTGCTGAGTTCCTTGGGCTTTCGATGCTCTTTCGATATCAAG 5158
Db 2810486 GTCATGTTCTGTTGGGCTGCTGAGTTCCTTGGGCTTTCGATGCTCTTTCGATATCAAG 2810427
QY 5159 TATGGCTTTACTTGGTTCGCGCAACGAGCAGTTCGATGCTGCTGCTGCTGCTGCTGCT 5218
Db 2810426 TATGGCTTTACTTGGTTCGCGCAACGAGCAGTTCGATGCTGCTGCTGCTGCTGCTGCT 2810367
QY 5219 GTGCTGCGAGAACGCAACGAGCAGTTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5278
Db 2810366 GTGCTGCGAGAACGCAACGAGCAGTTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2810307
QY 5279 ACCATCATCAGCAGCTTTCGCGTGAAGCTATCGCACTGAGCAGCTGCTGCTGCTGCTGCTGCT 5338
Db 2810306 ACCATCATCAGCAGCTTTCGCGTGAAGCTATCGCACTGAGCAGCTGCTGCTGCTGCTGCTGCT 2810247
QY 5339 ATGTTTCCAGCGAAAGCTTGGCTCAGGTTGGTTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 5398
Db 2810246 ATGTTTCCAGCGAAAGCTTGGCTCAGGTTGGTTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2810187
QY 5399 GTTTCACCATGAGCGAAAGCTTGGTGGCTTCCCATCTGGTTCAGCTTTCGAGTTCGAGTTC 5458
Db 2810186 GTTTCACCATGAGCGAAAGCTTGGTGGCTTCCCATCTGGTTCAGCTTTCGAGTTCGAGTTC 2810127
QY 5459 CGCACTAAGGCTGAGGATGTTTCCCAATGTGGATATCTTGTATGACATGTTGTTTCGACAC 5518
Db 2810126 CGCACTAAGGCTGAGGATGTTTCCCAATGTGGATATCTTGTATGACATGTTGTTTCGACAC 2810067

QY 5519 GTAAACCTCAACGCGACGCACTTTAACCCGCTGAAGAGCAGGCGGATGAAGTCAAGCA 5578
Db 2810066 GTAAACCTCAACGCGACGCACTTTAACCCGCTGAAGAGCAGGCGGATGAAGTCAAGCA 2810007
QY 5579 GGGGAGTGTGTGTGAATTCGATATTCATGATTCATTAAGCTGCGAGGTTATGAGGTAACC 5638
Db 2810006 GGGGAGTGTGTGTGAATTCGATATTCATGATTCATTAAGCTGCGAGGTTATGAGGTAACC 2809947
QY 5639 ACGCGGATGTGTGTTCGAATTCGAATTAAGAAACCGGACCTGTAACACATTCAGGTTGGGC 5698
Db 2809946 ACGCGGATGTGTGTTCGAATTCGAATTAAGAAACCGGACCTGTAACACATTCAGGTTGGGC 2809887
QY 5699 GAAATTGAAGCGGGAGGCAACCTGCTCAACGTCGCAAGAAAGAGCGGTGCCAGCAACA 5758
Db 2809886 GAAATTGAAGCGGGAGGCAACCTGCTCAACGTCGCAAGAAAGAGCGGTGCCAGCAACA 2809827
QY 5759 CCATAAGTTGAAACCTTGAAGTTCGACACAGGTTAGTAGAGGAGGACGTGACTCTACGC 5818
Db 2809826 CCATAAGTTGAAACCTTGAAGTTCGACACAGGTTAGTAGAGGAGGACGTGACTCTACGC 2809767
QY 5819 ATCTTTGACACCGGTACCGGTACCGTTCGAGATTTTAACTGTTCAACAGGTCATGCC 5878
Db 2809766 ATCTTTGACACCGGTACCGGTACCGTTCGAGATTTTAACTGTTCAACAGGTCATGCC 2809707
QY 5879 TCGGTGACCTGTGTGTGTCGACCCCGCAATCTTACCCACATTCGACATTCGTTCA 5938
Db 2809706 TCGGTGACCTGTGTGTGTCGACCCCGCAATCTTACCCACATTCGACATTCGTTCA 2809647
QY 5939 GCAGTAGCTTGTATATTTTCGCGCGCTG 5967
Db 2809646 GCAGTAGCTTGTATATTTTCGCGCGCTG 2809618

RESULT 2
US-09-738-626-2904
; Sequence 2904, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2904
; LENGTH: 1983
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2904

Query Match 31.4%; Score 1874.2; DB 10; Length 1983;
Best Local Similarity 96.6%; Pred.No. 0;
Matches 1915; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 3779 ATGGACCAATAGGACCTCGCGCAAGCATCTCTGCGCACATTCGCGGCAAGACAACATT 3838
|||||

APPLICANT:	TATEISHI, NAKO
APPLICANT:	SENOH, AKIHIRO
APPLICANT:	IKEDA, WASATO
APPLICANT:	OZAKI, AKIO
TITLE OF INVENTION:	NOVEL POLYNUCLEOTIDES
FILE REFERENCE:	249-125
CURRENT APPLICATION NUMBER:	US/09/738,626
CURRENT FILING DATE:	2000-12-18
PRIOR APPLICATION NUMBER:	JP 99/377484
PRIOR FILING DATE:	1999-12-16
PRIOR APPLICATION NUMBER:	JP 00/159162
PRIOR FILING DATE:	2000-04-07
PRIOR APPLICATION NUMBER:	JP 00/280988
PRIOR FILING DATE:	2000-08-03
NUMBER OF SEQ ID NOS:	7059
SOFTWARE:	PatentIn ver. 3.0
SEQ ID NO	2905
LENGTH:	1299
TYPE:	DNA
ORGANISM:	Corynebacterium glutamicum
US-09-738-626-2905	
Query Match	21.2%; Score 1267.8; DB 10; Length 1299;
Best Local Similarity	99.0%; Pred. No. 0;
Matches 1286;	Conservative 0; Mismatches 12; Indels 1; Gaps 1;
Qy	2327 TGTGTGGGGTATGCACACAGACTTTCCAGTTTGGCCCTGCGTACCATGTGACTCTCTC 2386
Db	2 TGTGTGGGGTATGCACACAGACTTTCCAGTTTGGCCCTGCGTACCATGTGACTCTCTC 61
Qy	2387 CGCAGGCGAGGCTCAATGATCCCAACGGGAATGTACGTGATGGAGATACCTCCACGTCT 2446
Db	62 CGCAGGCGAGGCTCAATGATCCCAACGGGAATGTACGTGATGGAGATACCTCCACGTCT 121
Qy	2447 ACTACCAGCAGATCCAGGTTTCCCTTCGCACCAAGCGCAGCGGTGGGCTCACACCA 2506
Db	122 ACTACCAGCAGATCCAGGTTTCCCTTCGCACCAAGCGCAGCGGTGGGCTCACACCA 181
Qy	2507 CCACGCGGTTGACGGACCGCAGCGATTCAGATGGACGACCTGCCGACGCTCTTTACC 2566
Db	182 CCACGCGGTTGACGGACCGCAGCGATTCAGATGGACGACCTGCCGACGCTCTTTACC 241
Qy	2567 CGGATGCATCTATGACCTGGATGCTATTCCGGTGGAGCGGTATTTACTGACGGCA 2626
Db	242 CGGATGCATCTATGACCTGGATGCTATTCCGGTGGAGCGGTATTTACTGACGGCA 301
Qy	2627 CACTTAACTTTTCTACACCGGCAACCTAAAAATTGACGAAAGCGCGCGCCACCCAAA 2686
Db	302 CACTTAACTTTTCTACACCGGCAACCTAAAAATTGACGAAAGCGCGCGCCACCCAAA 361
Qy	2687 ACCTGTGCAAGTCGAGGACCCAACTGGGCTGATGGCGCATTCATCGCGGTGCGCTA 2746
Db	362 ACCTGTGCAAGTCGAGGACCCAACTGGGCTGATGGCGCATTCATCGCGGTGCGCTA 421
Qy	2747 AAAATCGGTTATCGAGGACCGCCAGCGGTTTCACACCCCATTCACCGGATCCCATGA 2806
Db	422 AAAATCGGTTATCGAGGACCGCCAGCGGTTTCACACCCCATTCACCGGATCCCATGA 481
Qy	2807 TCAGCCCTGATGTGATGTTTGAACATGTTCTTGGGGCCCAACGGAACCTCACCG 2866
Db	482 TCAGCCCTGATGTGATGTTTGAACATGTTCTTGGGGCCCAACGGAACCTCACCG 541
Qy	2867 GTGAGCGGTTCTATACCGCTCGACAGATCTTGAAACTGGGAATTCCTCGGTGAAATCA 2926
Db	542 GTGAGCGGTTCTATACCGCTCGACAGATCTTGAAACTGGGAATTCCTCGGTGAAATCA 601
Qy	2927 CCTTTGACCTCAGTGTGCAACACCTGGTTCTCTCTGTATCTCTGATGCGGTGCGTACA 2986
Db	602 CCTTTGACCTCAGTGTGCAACACCTGGTTCTCTCTGTATCTCTGATGCGGTGCGTACA 661
Qy	2987 TGTGGGAATGCCCAACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGC 3046
Db	662 TGTGGGAATGCCCAACCTTTTACGCTTCGCGATGAAGAACTGCGGAAGATCTCGAGC 721

US-09-738-626-2907

Query Match	18.6%;	Score 1110.4;	DB 10;	Length 1152;	
Best Local Similarity	97.7%;	Pred. No. 0;			
Matches 1126;	Conservative	0;	Mismatches 26;	Indels 0;	Gaps 0;
QY	354	GTCAATTATCAAGAAATGCAAGTCAAGCTCAAGCTATATAAAATTTGAGGAAAGAAATGTTCCTCC	413		
DB	1	GTGCATTATCAAGAAATGCAGTCAAGCTTAAATAATTTGAAGGAAGAAATGTGTAC	60		
QY	414	CCCTCGGGGTGATTGATGGCTTTCCTCAACTCGAAACGGCATCATCAGGGAACCTCTCT	473		
DB	61	CCCCACGGGTGATTGATGGCTTTCCTCAACTCGAAACGGCATCATCAGGGAACCTCTCT	120		
QY	474	GGAGAACCAGCACTATAAACCGCAGGATTCACCCCGGAATCCCCACGATTTGTTCGCGT	533		
DB	121	GGAGAACCAGCACTATAAACCGCAGGATTCACCCCGGAATCCCCACGATTTGTTCGCGT	180		
QY	534	TTTATTGATCTTCATATCACGGTGGAAACGGTGGCGGTTCTCCTACGGGAACCGCAGGAC	593		
DB	181	TTTATTGATCTTCATATCACGGTGGAAACGGTGGCGGTTCTCCTACGGGAACCGCAGGAC	240		
QY	594	CAGCGAGGAACACCCGCGAGTATCATCCGGAACATGGCACAGCCGTGATTTGCCAAGC	653		
DB	241	CAGCGAGGAATGCCGCGAGTATCATCCGGAACATGGCACAGCCGTGATTTGCCAAGC	300		
QY	654	ATGGTTTCGCGCGGCTGACGCACATGGCAGCGACAGTGGAAACCTTATTCCTCTGTGT	713		
DB	301	ATGGTTTCGCGCGGCTGACGCACATGGCAGCGAGTGGAAACCTTATTCCTCTGTGT	360		
QY	714	GAAGAGTCTCTGTGTGCGCATTCACCTCGAGGGCCCTTTTCATCAACGATGCCGTGT	773		
DB	361	GAAGAGGGCTGTGTGCGCATTCACCTCGAGGGTCTTTTCATCAACGATGCCGTGT	420		
QY	774	GGTGCTCAAAACCGGATTCATTTTCCCGGCAACCCACAGATCTTGCCCGGTGATC	833		
DB	421	GGTGCTCAAAACCGGATTCATTTTCCCGGCAACCCACAGATCTTGCCCGGTGATC	480		
QY	834	CATCGGGAAAGGTTCGATCAAAATCGATCACAGTAGCGCGGAAACTGACATCTTTCT	893		
DB	481	CATCGGGAAAGGTTCGATCAAAATCGATCACAGTAGCGCGGAAACTGACATCTTTCT	540		
QY	894	GAGCTTCGATCTCTCGGAGCGCACCATCATCTTCTTCGCGGACACTGATGCA	953		
DB	541	GAGCTTCGATCTCTCGGAGCGCACCATCATCTTCTTCGCGGACACTGATGCA	600		
QY	954	GATTTGATACCACTACAGCGCAATTCCTTGCTAAAGAGAAAAATGTGAAGTCAAG	1013		
DB	601	GATTTGATACCACTACAGCGCAATTCCTTGCTAAAGAGAAAAATGTGAAGTCAAG	660		
QY	1014	GCTACGATTTGTTCAATGGATGCCCTCCGCTGCATCATAGGGCTCCGCGAGGTGGC	1073		
DB	561	GCTACGATTTGTTCAATGGATGCCCTCCGCTGCATCATAGGGTCCGCGAGGTGGC	720		
QY	1074	GCTTTGCTGCTGCGGCACGTGCGGGAGCATATGTGAGTTGATCGCGGAGCGGTG	1133		
DB	721	GCTTTGCTGCTGCGGCACGTGCGGGAGCATATGTGAGTTGATCGCGGAGCGGTG	780		
QY	1134	CATTTGCCGATGAACGGTCGATCTAGCTGTTCCACACGCGCTTTTCATCAGCGAC	1193		
DB	781	CATTTGCCGATGAACGGTCGATCTAGCTGTTCCACACGCGCTTTTCATCAGCGAC	840		
QY	1194	GCCATGAAGCGCGGAAATGCCAGCGGTGAGTACATTTTGGCGTTTGAAGCTCACC	1253		
DB	841	GCCATGAAGCGCGGAAATGCCAGCGGTGAGTACATTTTGGCGTTTGAAGCTCACC	900		
QY	1254	GTCACCGATGAGTCCCGCTCGGGATGGCGGCCCATTCGCGGGGACACGACACA	1313		
DB	901	GTCACCGATGAGTCCCGCTCGGGATGGCGGCCCATTCGCGGGGATACGACACA	960		
QY	1314	CTACGAGTCAGTTCTGTGCACCACTGCGCAGGGGTATGACGCTTATGACGCAACCTC	1373		
DB	961	CTACGAGTCAGTTCTGTGCACCACTGCGCAGGGGTATGACGCTTATGACGCAACCTC	1020		

Qy	1374	CACACCTCAACCGTCGCGCTAAATTCCTCGGTCTTGGCGNCTACAGAAATCGCTAAATCC	1433
Db	1021	CACACCTCAACCGTCGCGCGCAAAATTCCTGGACTTACGATCAGAAATCGTTAAATCC	1080
Qy	1434	AACCTCGCAATTTTGTGGTCTTTGACTCAAAACGGCAGGTCGCAATTTAGGT	1493
Db	1081	AACCTGTAAATTTTGTGGTCTTTGACTCAAAACGGCCAGTTACAAACAGTCCATTTAGAC	1140
Qy	1494	CATCAAGTACTT	1505
Db	1141	CATCAAGTAATT	1152

RESULT 5

```

1  RESULT 3
2  US-09-738-626-2906
3  ; sequence 2906, Application US/09738626
4  ; Publication No. US20020197605A1
5  ; GENERAL INFORMATION:
6  ; APPLICANT: NAKAGAWA, SATOSHI
7  ; APPLICANT: MISOGUCHI, HIROSHI
8  ; APPLICANT: ANDO, SEIKO
9  ; APPLICANT: HAYASHI, MIKIRO
10 ; APPLICANT: OCHIAI, KEIHO
11 ; APPLICANT: YOKOI, HARUHIKO
12 ; APPLICANT: TATEISHI, NAOKO
13 ; APPLICANT: SENOH, AKIHIRO
14 ; APPLICANT: IKEDA, MASATO
15 ; APPLICANT: OZAKI, AKIO
16 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
17 ; FILE REFERENCE: 249-125
18 ; CURRENT APPLICATION NUMBER: US/09/738,626
19 ; CURRENT FILING DATE: 2000-12-18
20 ; PRIOR APPLICATION NUMBER: JP 99/377484
21 ; PRIOR FILING DATE: 1999-12-16
22 ; PRIOR APPLICATION NUMBER: JP 00/159162
23 ; PRIOR FILING DATE: 2000-04-07
24 ; PRIOR APPLICATION NUMBER: JP 00/280988
25 ; PRIOR FILING DATE: 2000-08-03
26 ; NUMBER OF SEQ ID NOS: 7059
27 ; SOFTWARE: PatenIn ver. 3.0
28 ; SEQ ID NO 2906
29 ; LENGTH: 759
30 ; TYPE: DNA
31 ; ORGANISM: Corynebacterium glutamicum
32 US-09-738-626-2906

```

Query Match	11.3%	Score	671.8	DB	10	Length	759
Best Local Similarity	93.7%	Pred. No.	2.9e-201				
Matches	711	Conservative	0	Mismatches	47	Indels	1
Gaps	1						
QY	1548	ATGGAAATCACTATCTGCAAGACGACGAGAAAGTCGCAAGACGAGTTGCAGTCCTCAATC	1607				
Db							
QY	1	ATFGACATCATCTGCAAGACGACGAGAAAGTCGGCAAGACGACGCGACCTGATC	60				
Db							
QY	1608	GCACCCCTTCGCCAACAAGSGTGGAACTTTGGGGCTTGCAACAGAGATCCTCACCATGAGT	1667				
Db							
QY	61	GCACCCCTTCGCAACTAAGGGCGCAACCTTGGGGCTTGCAACTCGATCGTCACTTTGAC	120				
Db							
QY	1668	ACCTACCAAGAGCTCATTCGCATGTATGAGCTGGGAGTGTCTATCAAGAACTGCGAAG	1721				
Db							
QY	121	ACCTACCAAGAGCTCATTCGCATGTATGAGCTGGGAGTGTCTATTCAAGAACTGCGAAG	180				
Db							
QY	1728	GCATTCCTTTGGATGAATACGTGGGACTTAACCCGTGACGATGAAACAGCTACTTTAA	1781				
Db							
QY	181	GCATTCCTTTGGATGAATACGTGGGATTAAACGCGGACGATGAAACAGCTACTTCAA	240				
Db							
QY	1788	ACCATTCCGAAGAGTTTCACTGACACATCGACATCGTTGATGAGAGAGTCTACAGCCCA	1841				
Db							
QY	241	ACCATTCTGTAAGAGATTCACTGACCACATCGACATCGTTGATGAGAGAGTCTACAGCCCA	300				
Db							
QY	1848	GATGTTGCAAAACCCCTGATCCATACGAAGCAGCTGCAGAGTATGAGGCAAGATCGCTGCA	1901				
Db							

Db 301 GATGGTCAAAACCCCTGATCATACGAAGCAGCTGCAGAGTATGAGCAAGATCGCTGCA 360
QY 1908 GAATCGTTGAAGTTCAATTCCTTGGCATCGCGGAAACGG-CACATCCCTTTCATTGAA 1966
Db 361 GAATCGTTGATGTTTCAATTCCTTGGCATCGCGGAAACGGCCACATCGTTTCAATGAG 420
QY 1967 CCATCATCTTCTGTGTCAGAGTACGACAAAGGTCCAGGCGCTGCACCTAAAACCTGTGGAG 2026
Db 421 CCATCATCTTCTGTGTCAGAGTACGACAAAGGTCCAGGCGCTGCACCTAAAACCTGTGGAG 480
QY 2027 GACACGCTCGGATCTTTCACACCATCGAAGAGTCCCAACCCAGCCGCTCACCCAGGT 2086
Db 481 GACACGCTCGGATCTTTCACACCATCGAAGAGTCCCAACCCAGCCGCTCACCCAGGT 540
QY 2087 TTGGGCACATTGTCCCGCGCGCAAAACATCGTTGGTGGCAACTGGTCAAGGAAAGCC 2146
Db 541 TTGGGCACATTGTCCCGCGCGCAAAACATCGTTGGTGGCAACTGGTCAAGGAAAGCC 600
QY 2147 GAGCCATCGCGGGAAGTCTGTGGAAGCCGAGTACTGCTTCTGTGCCAGGTTCATCTCG 2206
Db 601 GAGCCATCGCGGGAAGTCTGTGGAAGCCGAGTACTGCTTCTGTGCCAGGTTCATCTCG 660
QY 2207 TAGATGCAACAATGCACCATCATCGTTGGATGAAGCAGCAGTATCCAAAGCTGGAAGAC 2266
Db 661 CAGATGCAACAATGCACCATCATCGTTGGATGAAGCAGCAGTATCCAAAGCTGGAAGAC 720
QY 2267 GCTGATCATACGCTCTCATGGAGCAATTAAGCTCGC 2305
Db 721 GCTGATCATACGCTCTCATGGAGCAATTAAGCTCGC 759

RESULT 6

US-09-746-660A-31/G
; Sequence 31, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CP2
; CURRENT APPLICATION NUMBER: US/09/746,660A
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: Patent In Vers. 2.0
; SEQ ID NO 31
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1036)
; OTHER INFORMATION: RXA00044

US-09-746-660A-31
Query Match 3.6%; Score 215.2; DB 11; Length 1059;
Best Local Similarity 96.5%; Pred. No. 1.7e-56;
Matches 220; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 AGTCCGTGAGCGCCACCATTTGATGTGGTGTACCGAGCTTGGGAGGCTTTTACATCT 60
Db 228 AGTCCGTGAGCGCCACCATTTGATGAGTGTGTCAACGAGCTTGGGAGGCTTTTACATCT 169
QY 61 AGCTCCCGTGGCGGTGGAGTGGGTTCATTACGGGTGGGATCACGCCGGTGAAGTTGGC 120
Db 168 ACATCGCGTGGCGGTGGAGTGGGTTCATTACGGGTGGGATCACGCCGGTGAAGTTGGC 109
QY 121 GAACCCATCTGTCTTCTTGTGGTGTGAGGGAACGAGTCCGGTGTGAGAGTGTTCAGATG 180
Db 108 GAAGCCATCTGTCTTCTTGTGGTGTGAGGGAACGAGTCCGGTGTGAGAGTGTTCAGATG 49
QY 181 TCTGCAGTTTTTAAGTATGATCATCATCAGCTTGGAAAGCTGAGTAAT 228
Db 48 TCTGCAGTTTTTAAGTATGATCATCATCAGCTTGGAAAGCTGAGTAAT 1

RESULT 7

US-09-974-300-829
; Sequence 829, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 829
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-829

Query Match 3.2%; Score 192; DB 10; Length 975;
Best Local Similarity 51.4%; Pred. No. 3.5e-49;
Matches 472; Conservative 0; Mismatches 440; Indels 5; Gaps 1;
QY 4240 TGTTCGTGAGATGATCAACCTGATGCGCATCTGCGCGTTCGCGTCTTTCGCGAGTGTGGT 4299
Db 27 TCTCGCTAATATGATTAACCTTGATCGCGGACTGCTTTCGATTTCTCCTCGGCTAAT 86
QY 4300 TGGTTTCAACCAACCAAGCTTTTCGGTGGCAATGAGTTCCTGGCGCCGCGCATTTGAT 4359
Db 87 CGGCTGGTGGCGGCTCAAGCGGTTTGGCGGAATCCGCTGCTCGCATCTGCTCGGCGT 146
QY 4360 GCGGATGTGTTCCTCAACCCCTGGTTAAGGGTACGACGTGGCGCCGACCATGACGCGG 4419
Db 147 CATGCTTGTGCACCCCTGATTTCGTTAACGCCCTGGGATACGCGGCTGCGGAGCAAGCGG 206
QY 4420 CGAATGCCAATGTGTCCTGCTTTCGGTGGTGGATTTGATGTTGCTCAAGTGTGTACCGGGC 4479
Db 207 AGAGATTCCTGCTGGAATTTATTCGGTTTGAAGTGCAGAGGTCGCGTACCGAGGCA 266
QY 4480 CGTGTCTCTGCTGTGGTGTCTTCTTGGATTCTGCAACGATCGAGAGTTCCTGCACAA 4539
Db 267 GGTGCTTCCCATTTTGCCTGCTTCTATTGTGCGCAAAATTTGACGCTTTTTCACAA 326
QY 4540 GCGACTCATGGGCACTGCAGACTTCTGATCACCCGAGTGTGACTCTGCTGCTCACC 4599

Db	327	CGGAGCGCTGAGAGCAATTCAGCTCCTCTGTTGGGCACCGATTACGCTGCTGTCCACAGG	386
Qy	4600	CTTCCTTACGTTTCATTGCTATTGGTCCAGCAATGCGTGGTGGGTGACTTIGCTGGCACA	4659
Db	387	TTTTCGCTTCATTATTGTCCTTCGACCGATTACATTTGCCATCGGAATCTTTTACATC	446
Qy	4660	CGGCTCGAGGGACTCTATGANTTCGGTGGTCCAGTCGCGGCTGCTTTTTCGGTCTGGT	4719
Db	447	AGCGCTTATTGCTGTCTTTGAACAATTCGCGGTACTGGCGGACTGTTTACGAGAGGCT	506
Qy	4720	CTACTCACCAAATCGTTATCACTGCTGTGCACCAAGTCTCTCCCGCCAAATTGAGCTGGAGCT	4779
Db	507	TTACGGCGGCTCGTCAATTACCGGAATGCATCATACATCTCTGCGGTGACCTTCAGCT	566
Qy	4780	GTTC-----AACGAGGTGGATCTCTTCATCTTCGCAACGCGATCCATGCCCAATATCGC	4833
Db	567	GATCGGCTCAAAGCTCGGCGAACATTTTATGGCGATGCTCGCGGTGTCCCAATATTGC	626
Qy	4834	GCAGGTGCGAGCATGTTTGGCATGTCTCTTCCTAGCGAAGAGTGAAAGCTCAAGGCGCT	4893
Db	627	ACAAGGTTACGCGCGCTTCGCGATGATGTTTATTGTCAAAGATGAGAAGCAGAAAGCGCT	686
Qy	4894	TGCAGGTGCTTCAGGTGCTCTCCGCTGTCTTGGTATTACAGAGCGCTCGCATCTTCGGTGT	4953
Db	687	GTCCCTGACATCCGGAATTCAGCTTATCTCGGAATTACCGAACCGGCATTGTTCGGAGT	746
Qy	4954	GAACCTTCGCTCGCTGGCGGCGTCTACATTGGTATCGGTACGCGAGCTATCGGTGGCGC	5013
Db	747	GAATTCAGATACAGATTTCCGTTTCGTGATCGCGATGATCAGTTCGSGGATTGCCGSTAT	806
Qy	5014	TTTGATTGCACTCTTTGATATCAAGCGAGTTGGCTTGGGCGCTGCAGGTTTCTTTGGTGT	5073
Db	807	GTTTCATTTCCTGCAAGGGTTTTTGGCAAGCTCTGTGCGGCTCGGCGCGCTACTGGAAT	866
Qy	5074	TGTTTCTATTGATGCTCCAGATATGGPCATGTTCTTCTGTTTGGCGGTAGTTACCTTTGT	5133
Db	867	TTTCTCTATTATGAGCCAACTACTGGGAGAGCGTTTCGATCGGAATGCGGATCGTGTTGAT	926
Qy	5134	CATCGCATTCGGGCGAC	5151
Db	927	CGCGCGGTTTGC CGGAAC	944

RESULT 8

```

US-08-781-986A-56
RESOL: 8
; Sequence 56, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446

```

Db 11644 TTGGTACAGGTATACATCTGTGTACATTTATATTCACATCAGGATGGCTGGCG 11703
Qy 4701 GTCTGCTTTTGGTCTGTCTACTACCAATCGTTATCACTGTCTGTGACCAAGTCTTCC 4760
Db 11704 GACCAATATGATTTGATATGACCACTTGTATACAGACTACACCAATATGTTT 11763
Qy 4761 CGCAATATGAGTGGAGCTGTT-----CAACAGGAGGTGATCTTCTTCGCAACGG 4814
Db 11764 TAGCAGTAGATTTCCAAATGATGGGTAGCAGCTTAGCGGTACGATTTATGCGCAATG 11823
Qy 4815 CATCCATGGCAATATCGCGAGGTGACAGATGTTTGGCAGTGTCTTCTTCAGGAAGA 4874
Db 11824 TTCCGATTTCCAAATTTGTCAGGCTCTGCAGCAATTTGGAGCATGTTTGTCTATAC 11883
Qy 4875 GT-----GAAAGCTCAAGGCTTTCAGGCTCTTCAGGTGTCCTCGCTGTTCTTG 4925
Db 11884 GTCTGTAATGTTTAAAGAAAGGCTTGGCAATTAACATCTTGTATTTCTGTATGTTAG 11943
Qy 4926 GTATTACAGAGCTGCGATCTTCGGTGTGAACCTTCGCTGCGCTGGCGTCTTACATG 4985
Db 11944 GTGTACTGAACAGCCATGTTTCGGTGTGAACCTTACCTCTGAAATATCCATTTATCGCTG 12003
Qy 4986 GTATCGGTACCGAGCTATCGTGGCGCTTGTATTG 5021
Db 12004 CGATATCAAGCTCTGTGTATTGGGGGCAATCGTTG 12039

RESULT 9

US-09-974-300-685
; Sequence 685, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 685
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-685

Query Match 2.9%; Score 174.2; DB 10; Length 1380;
Best Local Similarity 48.9%; Pred. No. 2e-43;
Matches 524; Conservative 0; Mismatches 608; Indels 45; Gaps 4;
Qy 3779 ATGCACCATAGGACCTCGCGACGATCCTCGCGACATTCGCGGAGCAACATT 3838
Db 1 ATGATTCATAAGAAATCGCTGTACGTTTATACAGCTTCTTGGCGAGAGATATGTC 60
Qy 3839 GTCCGCCCGGCACACTGTCAACCGGTTTACGCTGTCTCAAGACACCAAGGATGTG 3898
Db 61 ATCAGTCAGCTATTCGCGACAAAGGCTTCGTTGGTGATTAAAGGACGAAAAATC 120
Qy 3899 GATCCCAAACTCTGGATGATGATCCAGATCTGAAGGACCTTTGAACCTGGCGCATG 3958
Db 121 GACGAGGAAAGGTCGAAGAGCTGGAAGGTGTAAGAGGACCTTTTCAAGCTCTGGTCAA 180
Qy 3959 TTCCAGATCATCTCGCGCCAGCGGATGTGATCATGTTTTCAAAGAACTCGATGAGCA 4018
Db 181 TACCAGATATTTCGCGAACAGGCTTGTAAATAAGTGT--ATGAACGTTTGGCAAA 237
Qy 4019 ACCTCCAAAGACATCGCTGTGTCCACAGAGAGCTCAAGATGTTGTGGCTAACACGCC 4078

RESULT 10

Db 238 GAGCTGAATCTTGACCAAAAGGAACGGTTCAGCACAGCATGCCGCAAAAGGAAATG 297
Qy 4079 AACTGGTTTACCGCTGTGTGAAGTATTTGGCGACATTTTCGTCGCTGATTCCAATC 4138
Db 298 AATCCTTTAGCGCGATTTTGGAAAACGCTGTCAAATATTTTCGTGCAATCATCCGCGC 357
Qy 4139 TTGTTTGGTGGCGTCTGCATCGCTATCAACAATGTTGGTGGCGAGGATCTGTTC 4198
Db 358 ATTGTGCAAGCGGTTTATTAATGGGCTGTGGGAATGATGAAGGATTTTCAGTGGGTC 417
Qy 4199 GTCGCGCAATCACTGGTGGAGATGTTCCCTCAGATCAGCGGTGTGTGAGATGATCAAC 4258
Db 418 GATCCTGATGGCGGCTGTA-----TAAATCTTGTAT 450
Qy 4259 CTGATGGATCTGCGCGCTTCGCGTCTTCCAGTGTGTGGTGTTCACCGCAACCAAG 4318
Db 451 ATGTTTCTCGAGCGCGCATTTATATCTTCCGATCTGTGCGGTGTCAGTGTGCGGAA 510
Qy 4319 GCTTTGGTGGCAATGAGTCTCTCGCGCGCGCATGATGATGGGATGGTGTTCCTCAAC 4378
Db 511 GAGTTTGGCAGCAATCTTACTTGGGGCTGTACTTGGAGGAATTTTAAATTCACCGGAGC 570
Qy 4379 CTGGTTAAGGCTAGCAGCTGCGCCACCATGACCGGGCGGAAATGCCAATGTGTC 4438
Db 571 TTGTTCAATCTTGGGACTTGCAACTGCATCACCGGATG-----TGATGAC 618
Qy 4439 CTGTTTGGTGGATGTTGCTCAAGCTGTTTACAGGCAACCGTTCCTGTGTGGTG 4498
Db 619 TTTTACGGACTCAAAATCGAGATGTTTGGCTATCAAGGAACGCTGTTCCTCAATTTGCT 678
Qy 4499 GTCTCTTGGATCTCGCAACGATCGAGAAGTCTCTGCACAGGCACTCATGGCACTGCA 4558
Db 679 GCGGTATATGATGAGTAAATCGAAAACGCGTAAAGAATCGTTCCAAATCGGATC 738
Qy 4559 GACTTCCCTGATCACCCCACTGTGATCTGTGTCTCACCGGCTTCCTTACGTTTACCT 4618
Db 739 GATCTATAGTAACACCGTTTGTACGATTTTCAACGGGATTCATTGCTTTTATCGG 798
Qy 4619 ATTGTTCAGCAATCGCTGGTGGTGTGCTGTGGCACAGGCTCTCAGGACTCTAT 4678
Db 799 ATCGGTCTTTAGGAAGATGCTTGTACAGGATCACGAACATTTTAAATATGTGTAT 858
Qy 4679 GATTTCCGTGTCCAGTGGCGGCTGTGCTTTTCCGCTCTGTCTACTACCAATCTTATC 4738
Db 859 GACGTTCGCGGCTGTTCGCGACTGATTTTCGCGGGAACATATCTCGTGTATCTTA 918
Qy 4739 ACTGTCTGCACCACTCTCTCCGCAATTTAGCTGTGAGCTGTTTC---AACAGGTTGA 4795
Db 919 ACTGTGTCCACCACTTTCATCGATTGAGCGCGGCTTATCGCTGATATCGCAGG 978
Qy 4796 TCCTTCATCTTCGCAACGCACTCCATGGCAATATCGCGGAGTGCAGCATGTTTGCA 4855
Db 979 AACTACTCTCTCCGATCTGTGCTATGTCCAATGTAGCCAGGCGGGCAGGAAAGCC 1038
Qy 4856 GTGTTCTTCTTACGCAAGAGTGAAGGCTCAAGGCGCTTTCAGGTGTCTCAGTGTCTCC 4915
Db 1039 GTATTCTTTTGGCAAGGCTGCAAAACAAAGAAATTCCTTACCTGCCCTTCTCT 1098
Qy 4916 GCTGTCTTGTATTACAGAGCTCGGATCTTCTCGGTGTGAACCTTCGCTGGCTGGCGG 4975
Db 1099 TCTTTTCTCGGATTTACCGAAGCTGTTTTCGGGTGTCACTCTCGTCTACCGCGCGG 1158
Qy 4976 TTCTACATTTGATCTCGTACCAGCATCTCGGTGGCGCTTTCGATTCGATCTTTCAT 5035
Db 1159 TTTATCGCGCAATGCGCGGGGAGCGCTGGAGCTGCATACGTTGCTTTTATGATGTG 1218
Qy 5036 AAGGCAAGTTCGTTGGG 5052
Db 1219 GCAGCAACGCAACGG 1235

```
US-09-815-242-9369
; Sequence 9369, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9369
; LENGTH: 1884
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1884)
US-09-815-242-9369

Query Match          2.7%; Score 163.6; DB 9; Length 1884;
Best Local Similarity 47.0%; Pred. No. 5.7e-40;
Matches 866; Conservative 0; Mismatches 914; Indels 84; Gaps 9;

QY 3779 ATGGACCAATAGGACCTCGCGCAACGCGATCTCGCGACATTCGCGGCAAGACCAAT 3838
Db 1 ATGAACAATCAGAAATTCGAAAAAGTCATCGATCGCTTGGCGGACGTGAATATGTC 60

QY 3839 GTCGCGCGGCACACTGTGCAACGCGTTTACGCTCGTCTCAAGACACCAAGGATGTG 3898
Db 61 AATAGTGTGCCACTGTGCGACTGCTACGTGTCTGTCATGTCATGTCATGTCATGTCATGTC 120

QY 3899 GATCCCAAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3958
Db 121 AATAAAGAAGTGATGAGAAGTTCGAAAAAGTTCGAAAAAGTTCGAAAAAGTTCGAAAAAG 180

QY 3959 TTCCAGATCATGTCGCGCGCGGCGATGTGATGATGATGATGATGATGATGATGATGATG 4018
Db 181 TACCAAAATATCTTTGGTACAGGTACAGTTTACAAAATGATGATGATGATGATGATGATGATG 240

QY 4019 ACCTCCAAAGACATCGCTGTGTCCACAGAGCAGCTCAAAAGATGTTGTGGCTTAACAAGCC 4078
Db 241 GCCTTACCAA-----CATCATCTAAGGATGACATGAAAGCAGAAGTTGCTTAACAAGGG 294

QY 4079 ACTGGTTACGCGGCTGTGAGAGTATGCGGACATTTTCGTCGCCCTGATTCCTCAATC 4138
Db 295 AACTGGTTCCAAAGCTGCTATCGCTACTTTTCGGTGTGATGTTTCGGTTCCTCAATCCTCCAGTT 354

QY 4139 TTGGTGTGGCGGCTGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 4198
Db 355 ATCGTAGCCACAGGCTCTCTTCATGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 414
```


Db 2049 CACAAGGTGGGATACGATGCTACCGATGTTATTACCAGCGGTTATTGCACAAGCGGGG 2108
Qy 4845 CATGTTGGCAGTGTCTTCTAGGAGAGAGTGAAGAGCTCAAGGGCTTGCAGGTGCTT 4904
Db 2109 CTGTTAGCTGTCTTTTCTCAACAAAATGTGAACAAAAGGTTGGCTTGTCTT 2168
Qy 4905 CAGTGTCTCGCTGTTCTTGGTATTACAGAGCTGCGATCTCGCGTGTGAACCTTCGCC 4964
Db 2169 CAAGTATTACGACTATTTTGGAAATCTAGCAACCACTGATATGGCGTGACTTTACCAT 2228
Qy 4965 TGGCGTGGCGTTTACATTTGGTATCGGTACCGCAGATATCGGTGGCGCTTGTATTCAC 5024
Db 2229 TGAACAAAACCATTTATTGACGCTTGTATGCGTGGGTATCGGTGGTGCATTTGGCTA 2298
Qy 5025 TCTTGTATCAAGGACGCTTGGTGGCGCTGCGAGTTTCTGGGTGTTGTTCTATTG 5084
Db 2289 TGAATCAGCTGAACAACTTTACGTTTGGCTTGGTTAGTATGTTGAGCTTGGCTGGCITTA 2348
Qy 5085 ATGCTCCAGATATGCTATGTTCTTGGTTTCCGGGTAGTTACCTTTGTCATCCGATTCG 5144
Db 2349 TTCTTGCAGACAAAAGATACCTCCCGATGATTACTGTTGCAATTTGGTCCGGAATTG 2408
Qy 5145 GCGGAGCGATTGCTTATGCGCTTACTTGGTTGCGCGCAAGCGGAGCATTTGATCCAGATG 5204
Db 2409 CTTTATCATTTGGCTTGTCTTAAGTTTGTCTTACGTT-----TTGAAGATCAAC 2459
Qy 5205 CAACCGCTGCTCCAGTGCCTGAGAGACGACCAAGCGGAGAGAGAGACCCCGCAGAT 5264
Db 2460 CTAATCCAGAACCACTGAAAAACAGAAAGTGAAGATGTTGGCACTGTAATAA 2519
Qy 5265 TTTCAAAGGATTCCACCATTCACGCA---CCTTTGACCGGTGAAGCTATCCACTGA 5321
Db 2520 CGAATCAGAGACAAAATTTATTAGCACTCCACITTCAGGTGAAATTTTACCGTAG 2579
Qy 5322 CGAGCGTCAGGATGCCATGTTTGGCCAGCGGAAGCTTGGCTCAGGTGTTGCGATCGTCC 5381
Db 2580 AAAAAGTACAAGACCTGTTTGGCTTCCAGGTGCTTTAGGAAAAGGTTTGCATTTAGC 2639
Qy 5382 CCACCAAGGCGAGCTGTTTCCAGGTGAGCGGAAGATCGTGTGGCTTCCATCTG 5441
Db 2640 CGACTGAAGGMAACTGTATGACCCGAGATGTTGAAATCACCACATATTTCGACAG 2699
Qy 5442 GTCAGCTTCCAGTCCGCACTAAGGCTGAGGATGTTTCCAAATGTGGATATCTTGATGC 5501
Db 2700 GACATGCTTGGCTGTGACGACACAGAGG-----GGCTTGAATTTAATGC 2747
Qy 5502 ACATTTGTTTGCACACCTTAACCTCAACGCGCAGCACTTTAAACCGCTGAAGAGAGG 5561
Db 2748 ATATTGGATGATACGTCGAATTTAGATGTTAAAGGCTTTGAATTTACGTGAACAAAG 2807
Qy 5562 GCGATGAATCAAGCAGGAGCTGCTGTGTGAATTCGATTTGATGCCATTAAGGCTG 5621
Db 2808 GTGATCTGTTAAAAAGAGATTTGCTAGTTACTTTTGTATTTGCTGCCATTAAGAAG 2867
Qy 5622 CAGGTTATCAGTACCGCGCGATGTTGTTTGAATTAACA 5664
Db 2868 CTGTTATCCGCTAGTTACCGGATTTGGTAAACGAATACGAA 2910

RESULT 13
US-09-070-927A-664
; Sequence 664, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville

STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICANT: US/09/070, 927A
FILING DATE: 04-May-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 664:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 664:
US-09-070-927A-664

Query Match 2.0%; Score 122.2; DB 10; Length 567;
Best Local Similarity 55.9%; Pred. No. 3e-27;
Matches 271; Conservative 1; Mismatches 209; Indels 4; Gaps 2;

Qy 3824 GGCGAAGACACATGTCGCGCGGACACATGTCGAACGGTTTACGCTCGTCTCAAA 3883
Db 83 GGAAGATATATCTTATCGCTGTGTCACATGTCGACACGCTTGTGCTCGTACTAAAG 142
Qy 3884 GACACCAAGGATGTGGATCGCCAAAGTCTGGATGATGATCCAGATCTGAAGGACCTTT 3943
Db 143 GATAATCAAAAGTTAATCAAAAGCTTTGGATGAAACCCAAATGTAAGGTTACCTTT 202
Qy 3944 GAACTGCGCGCATGTTCCAGATCATGTCGCGCGGCGATGTGGATCATGTTTCAAA 4003
Db 203 AAAATTGATGTCAGTATCAGTTTATTTGGTGGTACGCTTATTTTGTCTATGAT 262
Qy 4004 GAACTGATGACGCAAC---CTCCAAAGACATCGTGTGTCACAGAGAGCTCAAGAT 4060
Db 263 GAACCTTATAAAAAAACTGGCCTTTCAGAGCTTCAACAGACGACTTGAACCAATAGTT 322
Qy 4061 GTTGTGGCTACACACGCAACTGTTTCAGCGCTGCTGTGAAGTATTTGGCGGACATTTTC 4120
Db 323 GATAAAAAATAAAATTCAAATTCGAATTAATTAAGCTATCTATCAGAAATTTTC 382
Qy 4121 GTCCCGCTGATTCCAATCTTGGTGGTGGCTGCTCATGCTATGCTATCAACATGTGTTG 4180
Db 383 GTACCAATTATCTCGCAGCTTGTGCAAGTGGTTATTAAKGGCTCTACGAAATTTTAA 442
Qy 4181 GTTCCGCGAGGATCTGTTCCGCTCCGCAATCACTGGTGGAGATGTTCCCTCAGTACAGGCT 4240
Db 443 ACTCCGCGGGTCTTTTGGACCAAAATCAATCAGAGAAAGTATCTGCAATTAAGGA 501
Qy 4241 GTTCTGAGATGATCAACCTGATGCGCATCTGCGCGCTTCGCTTCTTCCAGTGTGTTGTT 4300
Db 502 ATTTCTCTATGATGATCAGCTTATGTCAGCTGCACCATTTATGTTTTTACCATCTAGTT 561
Qy 4301 GGTTT 4305
||| |


```
Db 562 GGTAT 566
RESULT 14
US-10-156-761-3925
; Sequence 3925, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3925
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1143)
US-10-156-761-3925
Query Match 2.0%; Score 121.2; DB 14; Length 1143;
Best Local Similarity 49.0%; Pred. No. 1.1e-26;
Matches 395; Conservative 0; Mismatches 36; Indels 18; Gaps 2;
Qy 515 CCCACGATGTTCCCGGTTTATTGATCTTCAATACAGTGGGAAACGGTGGCGGTT 574
Db 138 CCACTGGCTGGTCCCGGTTCTGATCTGCACACACGCGGGGGCGGCGCTCTT 197
Qy 575 TCCTACGGACGACGACGAGGAGGATTCACCGGCAATACACCGGGAACATGGC 634
Db 198 CACCTCGGGACGCGGAGGAGTTCGCGGGGCTCCACACCCACCGCTGCACGGC 257
Qy 635 GACCGTATGTTGCCAAGCATGTTTTCGCGCGCGGTGAGCGACTGCGACGAGTGG 694
Db 258 CACCACCGTCTGCTCCACCGTACCGCGGAGATGAGTCTGCGCCACGCGCGG 317
Qy 695 AAACCTTATCCCTTGTGTGAAGAGTCTGCTGCGGCGATTCACCTCGAGGCGCT 754
Db 318 GCTGCTCCGAGTGGCCGAGGAGGATTCGCGGCGATCCACTCGAGGCGCGTT 377
Qy 755 CATCAACGCGATGCGGTTGTGTTGCTCAAAACCGGATTTTCATTTTCCCGCAAC 814
Db 378 CATCTCGCGTGGCGAAGGCGCCCACTCCGAGGAATCTGCGGAGACCGGACCGG 437
Qy 815 AGATCTTCCCGGTTGATCATCGGGAAGGTTGATCAATTCATACATAGCGGC 874
Db 438 CGAGGTCCGAAGTGTACGCGGCGCGCGGCGGGAAGATGTCACGCTGGCCAC 497
Qy 875 GGAACCTGACAATCTTCTGAGTTCCTGATCTGCGGAGCGACACCATATGCTTC 934
Db 498 CGAATCCCGGGCGGATCATCTCCGTACGCTGCTGCGGAGCAGCGGGTGTGCGG 557
Qy 935 CTTCGGGACACTGATGAGATTTTGATACACTACACGCGCAATTCCTTGGCTAAGA 994
Db 558 GATCGGCGACGCGACGCGAGTACGAGCAGAGCGGTGGAGGCCATCGACCGGGT 612
Qy 995 GAAATATGACGCTACGCGATTTGTTCAATGCGATGCTCCCGTGCATCATAG 1054
Db 613 -----CGGACGGTGGCCACGCACTGTTCACGCGATGCCCGCGCTCGGCCAC 662
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Query Match 2.0%; Score 121.2; DB 14; Length 9025608;
Best Local Similarity 49.0%; Pred. No. 8.4e-24;
Matches 395; Conservative 0; Mismatches 393; Indels 18; Gaps 2;
Qy 515 CCCACGATGTTCCCGGTTTATTGATCTTCAATACAGTGGGAAACGGTGGCGGTT 574
Db 4865055 CCACGTGGTGGTCCCGGCTTCGTCGATCTGCACCAACACGCGGCGGCGGCTCCTT 4865114
Qy 575 TCCTACGGGAACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 634
Db 4865115 CACCTCGGCGACCGCGGAGGAGGTTCTCGGGGCGCTCCACACCCACCGCTGCACGCGAC 4865174
Qy 635 GACCGTATGTTGCCAAGCATGTTTCGCGCGCGCTCAGCCACTGCGGAGCGAGTGG 694
Db 4865175 CACCACCGCTGCTCGGCTCCACCGTACCGGCGGAGATGAGCGTGGCGCCAGCGCGCGG 4865234
Qy 695 AAACCTTATTCCTTGTGTGAAGAGTCTGCTGTCGCGCATTCACCTCGAGGCGCTTT 754
Db 4865235 GCTGCTCTCCGAGTGGCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4865294
Qy 755 CATCAACGATGCGGTTGTGGTGTCTCAAAACCGGATTCATTTTCCCGCAACCCAC 814
```

```
Db 4865295 CATCTCGCGTGGCGAAGGCGCCACCTCCGAGGAAGTCTCGCGGACCCGCGCC 4865354
Qy 815 AGATCTTTCGCCGGGTGATCATCGGGGAAAGTTGGATCAAAATCATCATAGTAGCGCC 874
Db 4865355 CGAGTCCGCAAGCTGATCAGCGCGCGCGCGAGGCGGAGATGTCACGCTGGCCAC 4865414
Qy 875 GGAACCTGACAATCTTCTCAGCTTCTCGATCTCTCGGAGCGCACCATCATCTTCTTC 934
Db 4865415 CGAATCTCCGGGGGCGATCGCTCCGTCACCTCGTGGCGAGCACGGGTGATCGCCG 4865474
Qy 935 CTTGGGACACATGATCAGATTTGTATACACTACAGCGCAATTGCCCTTGGCTAAAGA 994
Db 4865475 GATCGGACACAGGACCGGACGTACGAGCAGCGGTGGAGGCCATCGACCGGGT---- 4865529
Qy 995 GAAATGTGACGCTACGCTACGCTACGCTTTGTTCAATGCGATGCTCCGCTGCATCATAG 1054
Db 4865530 -----GCGACGGTCCCGACGACCTGTTCAAGCGGATGCCCGGCTCGGCCACCG 4865579
Qy 1055 GGCTCCCGGACGCTGGCGCTTTGCTTGTGGGACACGTGCGGGGACGCATATGTTGA 1114
Db 4865580 CACGCCGGCGCGATCGCCGCCCTCTGTGAGGACGAGCGGATCACGCTCGAGCTGATCAA 4865639
Qy 1115 GTTATCGCCGACGCGCTGCTATTTGGC---CGATGGAAACGTCGATCTAGCTGTTCCAA 1171
Db 4865640 CGACGGTACGCTATGACCGCGCGCCCTCCAGCTGGCGTTCCATCACCGGGCGCCGG 4865699
Qy 1172 CAACGCCCTTTTTCATCAGGACGCCATGGAAGCGCGCGGAATGCCAGACGGTGAGTACAT 1231
Db 4865700 CCGGTCGCGTTTCATCAGGACGCGGATGGACGCGCGGGCTTCGGCGACGCGCGCTATCT 4865759
Qy 1232 TTTGGGGGTTTGAACGTACCGTCCCGATGAGTGGAGTGGCGCGCTGCGCGATGGCGCGC 1291
Db 4865760 GCTCGGCCCGCTGGAGGTTCAGGTCACTGAGGGCGTGGCGGGCTGATGAGGCGCGGTT 4865819
Qy 1292 CATCGCGGGGGACACGACACTAG 1317
Db 4865820 GATCGCGGGCTCGACGCTCAGCTGG 4865845
```

Search completed: October 1, 2003, 05:00:44
Job time : 968 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 04:43:20 ; Search time 34 Seconds
(without alignments)
1869.632 Million cell updates/sec

Title: US-10-019-284A-2
Perfect score: 3342
Sequence: 1 MDHKLQAQRIILRDIGEDNI.....IEAGANLNVAKEAVPATP 661
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1609.5	48.2	651	2 S44257	phosphotransferase
2	1478.5	44.2	664	1 B32243	phosphotransferase
3	1227	36.7	480	2 S39978	scrA protein - Sta
4	1178	35.2	480	2 D90038	PTS system, sucros
5	1038.5	32.9	627	2 H96951	fusion, PTS system
6	1033	32.7	479	2 F82432	PTS system, sucros
7	1069	32.0	479	2 J00781	sucrose uptake pro
8	1013.5	30.3	470	2 H83926	PTS system, trehal
9	999	29.9	470	2 C69725	phosphotransferase
10	990.5	29.6	632	2 S68599	phosphotransferase
11	960	28.7	627	2 F95200	PTS system IIAAC c
12	957	28.6	617	2 AB1167	phosphotransferase
13	956	28.6	627	2 E98067	phosphotransferase
14	949	28.4	655	2 C95220	trehalose PTS syst
15	934	27.9	705	2 A99084	phosphotransferase
16	933.5	27.9	475	2 C89813	phosphotransferase
17	930.5	27.8	630	2 H83686	hypothetical prote
18	929.5	27.8	636	2 C83724	PTS system, beta-g
19	927	27.7	633	2 AC1436	PTS system, beta-g
20	924.5	27.7	634	2 AD1078	PTS system, beta-g
21	901	27.0	609	2 I40406	beta-glucoside per
22	896	26.8	631	2 B42603	beta-glucoside-spe
23	887	26.5	609	2 T47097	hypothetical prote
24	860.5	25.7	628	2 D97073	PTS system, beta-g
25	859	25.7	617	2 AC1421	beta-glucoside-spe
26	829	24.8	618	2 AC1204	phosphotransferase
27	824	24.7	625	2 C25977	phosphotransferase
28	822.5	24.6	636	2 D86807	hypothetical prote
29	816	24.4	456	2 S62331	phosphotransferase

30	801.5	24.0	612	2	A97935	hypothetical prote
31	799.5	23.9	612	2	B95067	hypothetical prote
32	781	23.4	458	2	H83881	-PTS system, sucros
33	770.5	23.1	455	1	WOEBST	phosphotransferase
34	751.5	22.5	460	2	A39938	phosphotransferase
35	747.5	22.4	640	2	AB1423	beta-glucoside-spe
36	719	21.5	494	2	AG1231	PTS system trehalo
37	706	21.1	459	2	JU0293	levansucrase synth
38	703.5	21.1	494	2	AF1585	PTS system trehalo
39	669.5	20.0	483	2	A10449	protein-Npi-phosph
40	659	19.7	473	2	A98281	trehalose specific
41	655	19.6	473	2	C65236	phosphotransferase
42	652	19.5	473	2	A86122	PTS system enzyme
43	641	19.2	478	2	B82263	PTS system, trehal
44	630.5	18.9	372	2	I39868	sac operon regulat
45	620	18.6	472	2	AF1058	protein-Npi-phosph

ALIGNMENTS

RESULT 1

S44257

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose specific - *Pediococcus per*

C:Species: *Pediococcus pentosaceus*

C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999

C:Accession: S44257

R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.

submitted to the EMBL Data Library, April 1994

A:Description: The sucrose and raffinose operons of *Pediococcus pentosaceus* PPel.0.

A:Reference number: S44252

A:Accession: S44257

A:Molecule type: DNA

A:Residues: 1-651 <LEED>

A:Cross-references: EMBL:Z32771; NID:G493728; PIDN:CAA83668.1; PID:G475968

C:Genetics:

A:Gene: scrA

C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

C:Keywords: phosphotransferase

F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III hc

Query Match 48.2%; Score 1609.5; DB 2; Length 651;

Best Local Similarity 49.6%; Pred. No. 1.8e-101;

Matches 331; Conservative 114; Mismatches 196; Indels 27; Gaps 8;

QY	1	MDHKLQAQRIILRDIGEDNI	VAAAHCA	TRLRILV	KDTRD	RQSLDD	DDPDL	KGFTETG	GM	60										
DB	1	MNHQVADRVLNAI	-GKNNIQA	AHCA	TRLRILV	KDESKID	QALDD	DDADV	KGFTETG	59										
QY	61	FOITVPGD	VDVRFKELDD	ATSKD	IAVSTEQ	LKD	VYA--	NNANWFS	RAVKVLAD	IFVPLI	118									
DB	60	YQIITGPGD	VYDALIV	KTGLK	-EVPDD	IKAV	AAAGQ	NKNPLMD	FLKVLSD	IFIPIV	118									
QY	119	PLVGGGLM	LNINVLVAQ	DLFG	PSQSLV	MPQISG	VAEMIN	LMASAP	FAELP	VLVGFTA	178									
DB	119	PALVAGGLM	LNINVLVAQ	DLFG	PSQSLV	MPQISG	VAEMIN	LMASAP	FAELP	VLVGFTA	178									
QY	179	TRFSGN	FLGAGIG	MANVFT	LYNGYD	VAATW	AGEMP	MSLFG	LDVAQ	GYOGV	PLVP	238								
DB	179	TRFSGN	FLGAGIG	MANVFT	LYNGYD	VAATW	AGEMP	MSLFG	LDVAQ	GYOGV	PLVP	238								
QY	239	LVVSWIL	ATIEFK	HLKRLM	GTAD	FLITP	VLITL	TGTLT	FAIGP	AMRWG	DLAHL	AGSQ	298							
DB	239	LGVAIT	ATLEK	FFHHK	KGAFD	FTT	PMFAI	VTGTLT	FTIVG	PVLT	VS	DALTN	LVG	298						
QY	299	LYDFG	PGVGL	FLGV	LYSP	IVITG	LUHQ	SFP	PIELF	----	NOGS	FFIT	ATAS	MANIAQ	354					
DB	299	LYNSTG	WIGM	GIFGL	LYSAI	VTGL	HTQ	TPA	ETL	LANV	AKT	GGSF	FP	VAS	MANIQ	358				
QY	355	AACLA	VFEL	AKSE	KLK	LAG	AGSV	AVL	GIT	PE	ATFG	VNLR	WR	PEY	IGIT	AAIGG	ALI	414		
DB	359	AATL	FAIF	ATK	SQ	KOK	ALT	SS	AGV	SALL	GT	PE	ATFG	VNLR	WR	PEY	IGIT	AAIGG	ALI	418

QY 415 ALFDIKAVAGLGGVGVVSDAPDMVFLVCAVTVTFVAFGAIAIYGLYLVVRNGSIDP 474
 Db 419 GLFHLVSVAMGPAVSVIGFISIAKSKSPAFMLSAVSVFVAFPTFTFY-----AKRTIGDDR 474
 QY 475 DATAAPVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSVAIV 534
 Db 476 --KTAVDVFAAEAAVEAEIEVQ-EIPEEAASAANKAQVTDVLAAPLAGEAEVELSVN 532
 QY 519 DAMFASGKLGSVAIVPTKQGLVSPVSGKIVVAPSPGHAFVTRKAEDGNSVDILMHGIF 578
 Db 533 DPFVSSEAMGKIAIKPSGNTVYAPVDTGVAFTDTHAYGI--KSDNGA--EILIHIGI 588
 QY 579 DTVNLNGHFHPLKKQGDVDEKAGELLCEFDIDAKAAGYEVTTPIVSNYKTKTGPVNTY- 637
 Db 589 DTVSMEGKFEQKQVADQKIKKGDVGLGTFDSKIAEAGLDNTTTFIVTINTADYASVETLA 648
 QY 638 GLGEIEAGANLLNVAK 653
 Db 642 AASSVAAT 649

RESULT 2

B32243
 Phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus mutans
 C:Species: Streptococcus mutans
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: B32243
 R:Sato, Y.; Poy, F.; Jacobson, G.R.; Kuramitsu, H.K.
 J. Bacteriol. 171, 263-271, 1989

A:Title: Characterization and sequence analysis of the scrA gene encoding enzyme II (scr)
 A:Reference number: A32243; PMID:89123027; PMID:2536656
 A:Accession: B32243
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-664 <SAT>

A:Cross-references: GB:M22711; NID:g153799; PIDN:AAA26971.1; PID:g153801
 C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase
 C:Keywords: phosphoprotein, phosphotransferase
 F:531-664/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol

Query Match 44.2%; Score 1478.5; DB 1; Length 664;

Best Local Similarity 46.0%; Pred. No. 1.4e-92;

Matches 311; Conservative 127; Mismatches 203; Indels 35; Gaps 13;

QY 1 MDHKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLKDTKVDQSLDDDDPDLKGTETGGM 60
 Db 1 MDYSKVAEIVTAV-GKDNLVAAAHCAATRLRLVLKDDSKVDKALDNADVKGTFKTDQ 59
 QY 61 FOIIVGPGVDVHFKELDDATSKDIAVSTEQKLKDVVANA--NWFSRAVKVLADIFVPLI 118
 Db 60 QYVIIGPGVDVHFKELDDATSKDIAVSTEQKLKDVVANA--NWFSRAVKVLADIFVPLI 118
 QY 119 PILVGGGLLMAINNVLVAQDLFGQSLVEMFPOISGVAEMINLMASAPFAFLPVLVGFTA 178
 Db 119 PALVAGGLMALNNFLTSEGLFQTKSLVQOFPPIKSGSDMIQLMSAAPFELPVLVGISA 178
 QY 179 TKRFGGNEFLGAGIGMAMVFPPTLVNGYDVAA-----TMTAGEMP--MWSFLGLDVAQAY 231
 Db 179 AKREGANQFLGASIGMIVAPGAANIIGLAANAPISKAATIGAYCFWNIFGLHVTQAY 238
 QY 232 QGTIVPLVNVSWILATIEKFLHRLMGTADFLITPVLTLLGLFTFFIAIGPAMRWVGL 291
 Db 239 TYQIVPVLVAVLLSLIEKFFHRLPSAVDFETPLLSVIITGFTFFIVIGPVKREVSDW 298
 QY 292 LAHGLQGLYDGGPGVGLFLGLVYSPVITGLHQSFPPIELEI---FNQG---GSFIFAT 345
 Db 299 LTNGIYVWLYDTTGLMGVFGALYSPVMTGLHQSFPPIELEI---FNQG---GSFIFAT 358
 QY 346 ASMANAQGAACIAVFLAKSEKLGAGASGAVYLGITETPAIFGNVRLRWPYIGIG 405
 Db 359 ASMANVAQGAATEAIVFLTKDKMKKGLSSSGVSGVALLGITETPALFGVNLKRYEPFFCALI 418
 QY 406 TAAIGGALIALFDIKAVAGLGGVGVVSDAPDMVFLVCAVTVTFVAFGAIAIYGLYLV 465

RESULT 3

S39978

scrA protein - Staphylococcus xylosus

C:Species: Staphylococcus xylosus

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999

C:Accession: S39978

R:Wagner, E.; Goetz, F.; Brueckner, R.

Mol. Gen. Genet. 241, 33-41, 1993

A:Title: Cloning and characterization of the scrA gene encoding the sucrose-specific

A:Reference number: S39976; PMID:94049686; PMID:8232209

A:Accession: S39978

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-480 <WAG>

A:Cross-references: EMBL:X69800; NID:g407905; PIDN:CAA49461.1; PID:g407908

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 36.7%; Score 1227; DB 2; Length 480;

Best Local Similarity 52.7%; Pred. No. 1.1e-75;

Matches 252; Conservative 83; Mismatches 133; Indels 10; Gaps 6;

QY 1 MDHKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLKDTKVDQSLDDDDPDLKGTETGGM 60
 Db 1 MNYKSAENILQALGGEDNVEAMTHCATRLRLVLKDEGLVDEKALGDMVVKGFSTGGQ 60
 QY 61 FOIIVGPGVDVHFKELDDATSKDIAVSTEQKLKDVVANNAMFSAVKVLADIFVPLIPI 120
 Db 61 QYVIIGSGTVNKFSELEKITGKE--ASSVSEVKTQGTKNMNPQRFVKMLSDIFVPIIPA 119
 QY 121 LVGGGLLMAINNVLVAQDLF-GPOSLEVMFPQISGVAEMINLMASAPFAFLPVLVGFTAT 179
 Db 120 IVAGLLMGINNLTAPGIFYDNQSLIEVQNFSGLAEMINIFANAPFTLLPILIGSAA 179
 QY 180 KRFGNEFLGAGIGMAMVFPPTLVNGYDVAAATMTAG--EMPWMSFLGLDVAQAYOGTVLPV 238
 Db 180 KRFGGNAYLGAALGMLVHPBELMAYDYPKALEAGKEIPHNLFGLEINQVGYQGVLP 239
 QY 239 LVVSHILATIEKFLHRLMGTADFLITPVLTLLTGLTFTIAGPAMRWVGDILAHGLOG 298
 Db 240 LVATYILATIEKGRKRVPTVLDNLLTPLLAILSTGFTFSFGPLRTGLYWSLDGLTW 299
 QY 299 LYDFGPGVGLLFLGLVYSPVITGLHQSFPPIELEI---NOGGSFIFATASMANIAQ 354
 Db 300 LYDFGGAIGGLFLGLYAPIVITGNHHSFIAIEHQLIADSSSTSGSFIFPIATMSIAQ 359
 QY 355 AACLAIVFLAK--SEKLKGLAGASGAVYLGITETPAIFGNVRLRWPYIGITGAAGAL 413
 Db 360 AAALAAFFIIEKNKLLKGVASAAGVALLGITETPAMFGVNLKRYPIGIVGSGISAY 419
 QY 414 IALFDIKAVAGLGGVGVVSDAPDMVFLVCAVTVTFVAFGAIAIYGLYLVVRN 469
 Db 420 IAFKVKALALGTAGIPGFIISQNGNGWLVHGAMIIAFVAFGVYVYALSRYKYN 477

RESULT 4
D90038
PTS system, sucrose-specific IIBC component [imported] - Staphylococcus aureus (strain N
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: D90038
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogasawara, N.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A/Reference number: A89758; MUID:21311952; PMID:11418146
A/Accession: D90038
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-480 <KUR>
A/Cross-references: GB:BA000018; PID:g13702328; PIDN:BABA3469.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetics:
A/Gene: scrA
C/Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 35.2%; Score 1178; DB 2; Length 480;
Best Local Similarity 50.4%; Pred. No. 2.3e-72;
Matches 242; Conservative 94; Mismatches 130; Indels 14; Gaps 7;

Qy 1 MDHKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLTOKDVDRQSLDDDDPDLKGTFTETGM 60
Db 1 MNYKQSAEDILNAGIGGENLDMAHCAATRLRLVNDSELVNEEALNNMDVYKGTFTETGG 60

Qy 61 FQIIIGSGDGVHVFKELDLDDTSKDIAVSTEQKLDVANNWFSRAVKVLADIFVPLPI 120
Db 61 YQIIIGSGTVNKVFSELEKLTGRE-ASTTSEVRAQSAKNMPLQREVKMLSDIFVPLIPA 119

Qy 121 LVGGGLMALNNVLVAQDL-FGQSLVEMFPQISGVAEMINLMASAPPAPFLPVLTGFTAT 179
Db 120 IVAGGLMGNLNLITAKDLFEKSLIDVYSQFAGLAEMINVPANAPFTLLPLIFGSA 179

Qy 180 KRFEGNEFLGAGIGMAVFTLVNGYDVAATMTAGE-MPMWSLFGDLVDAQAGYQGTVP 238
Db 180 KRFEGNFFLGNLGMILVHLSLMSADFPKAVEAGRAIPYWDVFGHLNQVGGYQGLPM 239

Qy 239 LVVSWILATIEKFLHKLMTADPLTPVLTLLTGLTFTIAGPAMRWGDLHLAHGLQ 298
Db 240 LVAAYILASIEKGLRKVIPTVLDNLLTPLLSIFITAPLTFSPGPIRQLGYLSDGLTW 299

Qy 299 LYDFGPGVGLLGLVYSPVITGLHOSFPPIELF-NOGGSFIFATASMANIAQ 354
Db 300 LYFEGGALGLITGLLAPIVITGMHSHFIAVETTLIADATKGGSFIFPIATMSNVAQ 359

Qy 355 AACLAVFPLAK-SEKLGKLAGASVAVLQITEPAIFGVNLRWPPYIGTAAIGGAL 413
Db 360 GAATAAFPIIKQNKGLKGVASAGISALLGITEPAMFGVNLKURYPFPGAIVGSGISAY 419

Qy 414 IALFDIKAVAGLGAAGFLGVYSIDA-PDMVNFVCAVVTFFIAFGAAIAYGLVLRNGS 471
Db 420 IAPFKVRIALGTAGLPGFISINPVHAGWLHYFVGMTISFII-ATVTLILSKKAN 475

RESULT 5
H96951
fusion, PTS system, beta-glucosides specific IIBC component [imported] - Clostridium ad
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C/Accession: H96951
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4836, 2001
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: H96951
A/Status: preliminary

A/Molecule type: DNA
A/Residues: 1-627 <KUR>
A/Cross-references: GB:AR001437; PIDN:AAK78403.1; PID:g15023277; GSPDB:GN00168
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC0423
C/Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfere

Query Match 32.9%; Score 1098.5; DB 2; Length 627;
Best Local Similarity 36.8%; Pred. No. 7.7e-67;
Matches 236; Conservative 122; Mismatches 243; Indels 41; Gaps 8;

Qy 1 MDHKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLTOKDVDRQSLDDDDPDLKGTFTETGM 60
Db 1 MDNKTAKDILKLGGERKNWSAAHCAATRLRLVLTOKDVDRQSLDDDDPDLKGTFTETGM 60

Qy 61 FQIIIGSGDGVHVFKELDLDDTSKDIAVSTEQKLDVANNWFSRAVKVLADIFVPL 117
Db 61 YQIIIGSGTVNKVFSELEKLTGRE-ASTTSEVRAQSAKNMPLQREVKMLSDIFVPLIPA 116

Qy 118 IPIVGGGLMALNNVLVAQDL-FGQSLVEMFPQISGVAEMINLMASAPPAPFLPVLTGFT 177
Db 117 IPIAVASGLLGMGLGLDADFHLVNSK-----SGLYVILNMFNSNAFQFLPMIAFS 167

Qy 178 ATKRFEGNEFLGAGIGMAVFTLVNGYDVAATMTAGE-MPMWSLFGDLVDAQAGYQGTVP 237
Db 178 AAREFTNPLYAAALGAIMIHPLDQNAW---TLGEGIKHTINIFGLNIGMVYQGTVP 223

Qy 238 VLWVSWILATIEKFLHKLMTADPLTPVLTLLTGLTFTIAGPAMRWGDLHLAHGLQ 297
Db 224 ILISVWMSYIEKGLRKVIPTVLDNLLTPLLSIFITAPLTFSPGPIRQLGYLSDGLTW 283

Qy 298 GLYDFGPGVGLLGLVYSPVITGLHOSFPPIELFENQGG---SFIFATASMANIAQ 354
Db 284 TLYNITGFSGLVLEGLGLYSLIVITGHHSHFAIEAGLANPAHKNFLPLIWSMANVAQ 343

Qy 355 AACLAVFPLAK-SEKLGKLAGASVAVLQITEPAIFGVNLRWPPYIGTAAIGGAL 414
Db 344 GAALAVYFETROKMKSTAAPASFCLLIGTEPAIFGVNLRWPPYIGTAAIGGALGGY 403

Qy 415 ALFDIKAVAGLGAAGFLGVYSIDA-PDMVNFVCAVVTFFIAFGAAIAYGLVLRNGSIDP 474
Db 404 VFTKVAMTAVGTGPIGIAIVKGGSLNY---IIAMILAFGGAFIIMVLGKKEITEE 459

Qy 475 DATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEATALLSVSDAMFASGLSGVAIV 534
Db 460 DLNKETVNDKIKVEEVES-----VSPVNGKVLKKNVPDKTFAEGLIGDGV 509

Qy 535 PTKGQLVSPVSGKIVVAPPSGHAFVTRKAEKDSNVVDILMHIGFDTVNLNGTHFNPLKQ 594
Db 510 PEDGEVVSIDGTVVHVFETKHAIAKSK---NGVEMLIHIGDITVKMEGNGEKSFIN 565

Qy 595 GDEVKAGELLCEFDIDAIKAAGYEVTTPIVWSNYKKTGPVNT 636
Db 566 GEEVKGDKLQFDLDLVKEKAVSPVILVTIVTNHEDMGFVNS 607

RESULT 6
F82432
PTS system, sucrose-specific IIBC component VCA0653 [imported] - Vibrio cholerae (str
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: F82432
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qiu, H.; Dragoi, I.; Sellers
l, R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A>Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: F82432
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-479 <HEI>

A:Cross-references: GB:AE004395; GB:AE003853; NID:g9658068; PIDN:AAF96554.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0653

A:Map position: 2

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 32.7%; Score 1093; DB 2; Length 479;

Best Local Similarity 47.8%; Pred. No. 1.3e-66;

Matches 222; Conservative 85; Mismatches 15; Indels 6; Gaps 3;

QY 1 MDHDLAQRILRDIGGEDNIVAAHCAATRLRLVLTQKDVDRQSLDDDDPDLKGTFTTGM 60

DB 1 MDYPIAKQLLESGLGKSNIOALAHCAATRLRLVLTQKDVDRQSLDDDDPDLKGTFTTGM 60

QY 61 FOIIVPGDGVHVFKELDATSKDIAVSTEQKDVVANNANWFSRAVKVLADIIVPLPI 120

DB 61 YOLIFSGGIVNOVYAEMAKLGL-VEMSTNDVASAGAEKQNAQRAVRLGSLDIFVRIIPA 119

QY 121 LVGGGLMAINNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTAT 179

DB 120 IVAGGLMLGNLVTAPGLFIEGQSLIDANPGLADLASMINTFANAPFVLPVLLAFSAS 179

QY 180 KFGNEFLGAGICMAMVFTLVNGDYDAATMAGEMPWMSLFGLDVAQAGYOGTVLPVL 239

DB 180 RXFGNPFVGAALGMLVHPDNLNGWFGSGASVGTPTNVLGFEIEKVGQGSVLPVL 239

QY 240 VVSWILATIEKFLHRLMGADFLITPVLTLLTGLTFIAIGPAMRWVGLDLAHLQGL 299

DB 240 VSAYILAKIENGLRKIVPSVDNLLTPMLAIFITGLFTVTVGPTROVGFMLGDALNWL 299

QY 300 YDFGPGVGLLGLVYSPVITGLHQSPPIIELELF-NOGGSFIATASMANIAOGA 355

DB 300 YDSAGFVGALGFYAPFVINGMHSHFATETQLLADIVTGGTFFPIPAAMSNIAOGA 359

QY 356 ACLAVFFLAKSEKLAGAGSVAVLGTETPAIFGVNLRWPFYIGITAAIGGALIA 415

DB 360 AALAVGVMTKELKGVAPSGVALLGITETPAFEGVNLKRLPYPIAICGAALASAFIT 419

QY 416 LFDIKAVAGAGFLGVVSIDAPDMVFLVCA-VVTFVIAFGAAIAYGL 459

DB 420 LFNVAQALGAAGLPGIISINPQQIGYINGMAISFVAFAALTIV 463

RESULT 7

QJ0781

sucrose uptake protein - Vibrio alginolyticus

N:Alternate names: enzyme II-sucrose protein

C:Species: Vibrio alginolyticus

C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 18-Jun-1999

C:Accession: JQ0781

R:Blatch, G.L.; Scholle, R.R.; Woods, D.R.

Gene 95, 17-23, 1990

A:Title: Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose uptake-end

A:Reference number: JQ0781; MUID:91071601; PMID:2174811

A:Accession: JQ0781

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <BLA>

A:Cross-references: GB:M76768; GB:M30194; NID:g155261; PIDN:AAA27555.1; PID:g155262

C:Genetics:

A:Gene: scrA

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 32.0%; Score 1069; DB 2; Length 479;

Best Local Similarity 46.1%; Pred. No. 5.6e-65;

Matches 216; Conservative 94; Mismatches 151; Indels 8; Gaps 5;

QY 1 MDHDLAQRILRDIGGEDNIVAAHCAATRLRLVLTQKDVDRQSLDDDDPDLKGTFTTGM 60

DB 1 MNPYAVAKELLTLGGKSNITAHARATRLRLVADEQKIDQADINLEGVKQGFVAGQ 60

QY 61 FOIIVPGDGVHVFKELDATSKDIAVSTEQKDVVANNANWFSRAVKVLADIIVPLPI 120

DB 61 YOLIFSGGIVNOVYAEMAKLGTMS-EMSTNDVASAGAEKQNIYQPAVKGLSDIFVPIIPA 119

QY 121 LVGGGLMAINNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTAT 179

DB 120 IVAGGLMLGNLVTAPGLFIDGKSLTEANPGTLDANMINTFANAPFVLPVLLAFSAS 179

QY 180 KFGNEFLGAGICMAMVFTLVNGDYDAATMAGEMPWMSLFGLDVAQAGYOGTVLPVL 239

DB 180 KFGGNPYLGAALGMLVHPDNLNGWFGSGASVGNIPVNNILGFEIKVGQGSVLPVL 239

QY 240 VVSWILATIEKFLHRLMGADFLITPVLTLLTGLTFIAIGPAMRWVGLDLAHLQGL 299

DB 240 VSAYILAKIENGLRKIVPSVDNLLTPMLAIFITGLFTVTVGPTROVGFMLGDALNWL 299

QY 300 YDFGPGVGLLGLVYSPVITGLHQSPPIIELELF-NOGGSFIATASMANIAOGA 355

DB 300 YNTAGFVGAVGGLIYAPFVITGMHSHFATETQLLADIVTGGTFFPIPAAMSNVSOQA 359

QY 356 ACLAVFFLAKSEKLAGAGSVAVLGTETPAIFGVNLRWPFYIGITAAIGGALIA 415

DB 360 AALAVGVMTKELKGVAPSGVALLGITETPAFEGVNLKRLPYPIAICGAALASAFIT 419

QY 416 LFDIKAVAGAGFLGVVSIDAPDMVFLVCA-VVTFVIAFGAAIAYGL 463

DB 420 MFNVKAQALGAAGLPGIISI-TPDKIGYTAGVMVIAFLTAFLVTLVGI 467

RESULT 8

H83926

PTS system, trehalose-specific enzyme II, BC component BH2216 [imported] - Bacillus h

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: H83926

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: H83926

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-470 <STO>

A:Cross-references: GB:AF001514; GB:BA000004; NID:g10174613; PIDN:BA05935.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2216

C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

Query Match 30.3%; Score 1013.5; DB 2; Length 470;

Best Local Similarity 44.4%; Pred. No. 3.2e-61;

Matches 204; Conservative 90; Mismatches 150; Indels 15; Gaps 4;

QY 3 HKDLAQRILRDIGGEDNIVAAHCAATRLRLVLTQKDVDRQSLDDDDPDLKGTFTTGMFQ 62

DB 4 YKKEVNAIVEAGADNIQTATHCVTRLRFVLRDEGVQDKLESLDIDKGSFSTNGQFQ 63

QY 63 IIVGPGDGVHVFKELEL-DDATSKDIAVSTEQKDVVANNANWFSRAVKVLADIIVP 116

DB 64 VIIOGQTVDKYKELVAETGTEATKEDV-----KDAAKNTNVFQRAVKVLADIIVP 116

QY 117 LPIILVGGLLMAINNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVG 175

DB 117 ILPAIVTAGLLMGINNLTAGSIFVDGASVIDIHPQWADFAGIINLIANTAFVLPGLIG 176

QY 176 FTATKRFNGNEFLGAGICMAMVFTLVNGDYDAATMAGEMPWMSLFGLDVAQAGYOGTV 235

DB 177 VSAVRFSGSELLGVGLMLVHPDNLNGWFGSGAQLEGEIPTWNLFGLTIEQVYOGQV 236

QY 236 LPVLVVSILATIEKFLHRLMGADFLITPVLTLLTGLTFIAIGPAMRWVGLDLAHLQ 295

DB 237 LPILFSAWLAKIEFLRKRPVDSQLLVAPVALLIITGFAFAIGITITIGNITNV 296

QY 296 LOGLYDFGPGVGLLGLVYSPVITGLHQSPPIIELELFNO-GGSFIATASMANIAOG 354

Db 297 FTSIFAAPLVGGLYGLIAPLVVGMHTFLAVDLQLIGTIGTFFLPILVLSIAQG 356
 Qy 355 AACLAVFFLAKSKLGLAGASVAVLIGTEPAIGVNLRLRWPFYIGIGTAAIGGALI 414
 Db 357 SAALAMWFAVRDEKLGSLSSAVSAVYLGTEPAMFGVNRKFPVCAIIISAAIGGAFI 416
 Qy 415 ALFDIKAVAGLGAAGFLGVWSIDAPDMVMFLVCAVTVFI 453
 Db 417 TVNGVLANSIGVGLGPIRFSIQAGVGVFFIGMVIAPIL 455

RESULT 9
 C69725
 phosphotransferase system enzyme II (EC 2.7.1.69) phosphoenolpyruvate-dependent, trehalose
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: C69725; S67929; JG5037; I40497; S67864
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertoni
 C.; Bron, S.; Brouillette, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toononi, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: 469580; MUID:98044033; PMID:9384377
 A:Accession: C69725
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-470 <KUN>
 A:Cross-references: GB:Z99108; GB:AL009136; NID:g2633055; PIDN:CAB12609.1; PID:g2633104
 A:Experimental source: strain 168
 R:Helfert, C.; Gotsche, S.; Dahl, M.K.
 Mol. Microbiol. 16, 111-120, 1995
 A:Title: Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by a phospho-
 A:Reference number: 140497; MUID:95379486; PMID:7651129
 A:Accession: S67929
 A:Molecule type: DNA
 A:Residues: 324-362, 'L', 364-464, 'G', 466-470 <HEL>
 A:Cross-references: EMBL:X80203; NID:g580941
 R:Schoeck, F.; Dahl, M.K.
 Gene 175, 59-63, 1996
 A:Title: Analysis of DNA flanking the treA gene of Bacillus subtilis reveals genes encod
 A:Reference number: JG5037; MUID:97074649; PMID:8917076
 A:Accession: JG5037
 A:Molecule type: DNA
 A:Residues: 1-139, 'S', 141-362, 'L', 364-464, 'G', 466-470 <SC2>
 A:Cross-references: EMBL:Z54245; NID:gl000450; PIDN:CA91014.1; PID:g1000451
 C:Comment: This enzyme functions as the specific trehalose transporter. It belongs to th
 C:Genetics:
 A:Gene: treP; treB
 C:Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II
 C:Keywords: phosphotransferase; sugar transport system
 F:113-131/Domain: transmembrane #status predicted <TM1>
 F:160-181/Domain: transmembrane #status predicted <TM2>
 F:184-203/Domain: transmembrane #status predicted <TM3>
 F:230-245/Domain: transmembrane #status predicted <TM4>
 F:263-286/Domain: transmembrane #status predicted <TM5>
 F:305-325/Domain: transmembrane #status predicted <TM6>
 F:375-395/Domain: transmembrane #status predicted <TM7>
 F:402-421/Domain: transmembrane #status predicted <TM8>

Query Match 29.9%; Score 999; DB 2; Length 470;
 Best Local Similarity 43.9%; Pred. No. 3.1e-60;
 Matches 203; Conservative 93; Mismatches 160; Indels 6; Gaps 4;

RESULT 10

S68599
 phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - Streptococcus
 N:Alternate names: sucrose-specific enzyme II
 C:Species: Streptococcus sobrinus
 A:Variety: strain 6715
 C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
 C:Accession: S68599
 R:Chen, Y.Y.M.; Lee, L.N.; LeBlanc, D.J.
 Infect. Immun. 61, 2602-2610, 1993
 A:Title: Sequence analysis of scrA and scrB from Streptococcus sobrinus 6715.
 A:Reference number: S68598; MUID:93273516; PMID:8500898
 A:Accession: S68599
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-632 <CHE>
 A:Cross-references: EMBL:L06791
 C:Genetics:
 A:Gene: scrA
 C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfer
 C:Keywords: phosphotransferase; sugar transport system
 F:480-632/Domain: phosphotransferase system glucose-specific enzyme II, factor III h

Query Match 29.6%; Score 990.5; DB 2; Length 632;
 Best Local Similarity 35.5%; Pred. No. 1.7e-59;
 Matches 236; Conservative 114; Mismatches 270; Indels 45; Gaps 12;

Qy 1 MDHKDLAQRILRDIIGEDNIVAAAHCAIRLRLVKDTKDVRDQSLDDDDPDLKGTFTGGFQIIVG 66
 Db 1 MDNKQIAKEVIEALGGRDNRVSRVAHCATRLRVVVDEAKIDKRAENIDRVKGAFFNSGQ 60
 Qy 61 FQIIVGPDVDHVFKELD--ATSKDIIVSTQKLDVVANNANWFSRAVKVLADIFVPL 117
 Db 61 YQIIFGTGTVNKIYDEVYDLGLPTS-----STGEQKQEAQAQGNFMQRMTGDFVPI 115
 Qy 118 IPILVGGGLLMAINNVLAQDLFGPQSLVEMFPGQISGVAEMIN-----LMASAPFAF 169
 Db 116 IPIVLAIGLPMGLRGLTNTDTFLG-----FFGASSKIDINAFILYQVLTDTAF 166

170 LPVLVGTATKRFNGNEFLGAGIGMAMVPTLVNGYDVAAWTAGEMPWMSLFGDLVQAQ 229
167 LPALLIASFKVFGNGPVLGVLGMLNPNALPNAYAVA---SGDAKALTFPGF-IEVY 221
230 GVGQVLPVLVSVTLATATEKFLHRLKMGTDALFTPTVLTLLTGFLTFIAIGPAMRWVG 289
222 GVGQVLPVAFVGMGTARLENLHRRVPEALDILLTPFLTFLVMSILGLFAIGPVFHSVE 281
290 DLLAHGLQGLYDFGGVGGLLFGLVSPVITGLHOSPPPIELFLFNOGGSFIF-ATASM 348
282 TVVLAATEWILALPFGIAGIIIGGLQQVIVTGVHFIENFLETQTLAETKANPFPNLLSA 341
349 ANIAOGAAGIAYFFLAKSKLGLAGAGSVALGTEPAIFGVNLRWRPFIIGITAA 408
342 ATAGQGVAVLAVAKTSKALKALYPSALSALGITEPAIFGVNLRKPKPVMGLVGS 401
409 IGGALIALFDIKAVLAGAGFGLVSVISIDAPDMV-----MFLVCAVVFVIFAGAAIYGLYLVR 468
402 AGGFIAALVGLKATGMSVTVLPGLLLFLNSQMPMTVTSITVACAIAF--ALTY--YFGYA 457
469 NSGIDPDATAPVPGCTTKAEAPAEFNSDSTIIQAPLTGAEALSSVSDAMFASGKL 528
458 DKEEDUSAKKPEAPAAAPVAETKSE-----VIASPLDGEAVELSKVNDPVFSEAMG 511
529 SGVAIVPTKQLVSPVSGKIIVAFPPSGHAFVARTKAEDGNSVDILMHIGFDTVNLNGTHF 588
512 KGIAPKPSGNTVSPVNGTVQIAFETGHAYGL--KSDNGA--EVLHVIGIDTVSMNGTGF 567
589 NPLKKQGDVEKAGELCEFDIDAIAKAGYEVTPPIVSVNKKTPGNTYVGLGEIAGANL 648
568 DQKVAANQTVKVGDLVGTDFDSAKIAEAGLDDTTWVITITADYSEYKPLAAGLAHAGDL 627
649 LNVAK 653
628 LELNK 632

RESULT 11

F95200
PTS system IIABC components [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95200
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-627 <KUR>
A:Cross-references: GB:AE005672; PID:AAK75799.1; PID:g14973217; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase

Query Match 28.7%; Score 960; DB 2; Length 627;
Best Local Similarity 33.3%; Pred. No. 1.9e-57;
Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;

QY 1 MDHKLDAQRIIRDIGEDNIVAAACATRLRLVLKDTKDVDRSLDDDDPKLGTFTGGM 60
DB 1 MNNEIAKKVDAIGGRENVSVAHACATRLRVYVDEKINVENLKVQGAFFNSGQ 60
QY 61 FQIIVGPDVHVFKE---LDDATSKDIAYSTEQDKOVVANNWFSRAVKVLADIPVL 117
DB 61 YQIIFGTGNKMYDEVVVLGLPTS-----SKDDMKAEVAKQGNWFORAITEGDVFPVI 115
QY 118 IPILVGGLLMAINNVLVAQDLFGPQSILVEMFPQISGVAEMINLMASAPFAFLPVLYGET 177

116 IPVIVATGLPMVGRGLENALEM-----PLPGDFATYTTQLLDTAFIILPGLVMS 165
178 ATRFEGNEFLGAGIGMAMVPTLVNGYDVAAWTAGEMPWMSLFGDLVQAQYQGTVLP 237
166 TFRVFGGNPAVIGIVLGMNLSGSLPNAYAVA---QGGEVTAMNFFGF-IPVVGLOGSVLP 221
238 VLVSWSILATATEKFLHRLKMGTDALFTPTVLTLLTGFLTFIAIGPAMRWVGDLLAHGLQ 297
222 AFIIGVGAKEKAVRVKVPDVIDLLVTPFVLLVMSILGLFVIGPVFHVENVILATK 281
298 GLYDFGGVGGLLFGLVSPVITGLHOSPPPIELFLFNOGGSFIF-ATASMANIAGAA 356
282 AILSNPFGGLGFLIGGVHQLIVSGVHHIFNLLEVLQVLAADHANPENAIIATAAQA 341
357 CLAVFFLAKSKLGLAGAGSVALGTEPAIFGVNLRWRPFIIGITAAIAGGALIAL 416
342 TVAVGVTKNPKLTLFAFPALSAFLGITEPAIFGVNLRKPKPFLSLIAGAIGGLASI 401
417 FDIKAVALGAAGFLGVSVISIDAPDMV-----MFLVCAVVFVIFAGAAIYGLYLVR 468
402 -----LGLAGTNGITIIPTGLYVGNGLPOLYLLMVAVSFALGFALTYMEFY----- 449
469 NSGIDPDATAPVPGCTTKAEAPAEFNSDSTIIQAPLTGAEALSSVSDAMFASGKL 527
450 --EDEVDTAAAKRAEVAEREVEVAPALQNETLV--TPIVGDVVALADNDPVFSSGAM 505
528 GSGVAIVPTKQLVSPVSGKIIVAFPPSGHAFVARTKAEDGNSVDILMHIGFDTVNLNGTH 587
506 GQGIIVKPSQGVVYVADAIEVSIAPFTGHAFGLKTR-----NGAEVLIHVIGIDTVSMNGDG 561
588 ENPLKKQGDVEKAGELCEFDIDAIAKAGYEVTPPIVSVN--YKKTGPVNTYGLGTEA 644
562 FETKVAQGNKVRAGDVLGTFDSNKIAAAGLDDTTWVITVNTGDIASVAPVAT---GSVAK 618
645 GANLLN 651
619 GDAVIEV 625

RESULT 12

AB1167
phosphotransferase system (PTS) beta-glucoside-specific enzyme IIABC component homolo
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AB1167
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Mehla
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Mehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1167
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-617 <GLA>
A:Cross-references: GB:NC_003210; PID:CA98816.1; PID:g16410127; GSPDB:GN00177
A:Experimental source: strain EGD-e

Query Match 28.6%; Score 957; DB 2; Length 617;
Best Local Similarity 35.1%; Pred. No. 3e-57;
Matches 225; Conservative 124; Mismatches 228; Indels 64; Gaps 17;

QY 1 MDHKLDAQRIIRDIGEDNIVAAACATRLRLVLKDTKDVDRSLDDDDPKLGTFTGGM 60
DB 1 MDYQIAKELLANVGGENVRVSVHACATRLRKLKVKKEKADKKQIESISGVLSVVENAQ 60
QY 61 FQIIVG--PGVDHV---FKELDDATSKDIAYSTEQDKOVVANNWFSRAVKVLADIPV 115

Db 61 LOVIIGNVGVYKALGSFTKLTDDGSEIAKGT---KD-----SGDNFLSKAIDVISGFT 114
QY 116 PLIPILVGGGLMANNVLAODLFGPOSIVEMFQISGVAEMINLMASAPFAFLPVLVG 175
Db 115 PILGALAGGMLKGLMLT---FG-----WLTSSGTQIILYAAADSVEYFLPLLA 165
QY 176 FTATKRFNGNEFLGAGIGMAVFPVLVNGYDVAATAGEMPMSLFLGLDVAQAGYQGTV 235
Db 166 YTAARKFGANPPVATAAGALVYPTMINLNFEGAHITELQIP-----VVLMSYSPSV 217
QY 236 LPVLVSMILATIEKFLHKRLMGTDFLITPVLTLTLGFTLFTIAIGPAMRWGDLIAHG 295
Db 218 IPILAVWFLSILERFLANSKIEHAATKELTPMICLMLIVPLTFLAFGLGTFIISQLASG 277
QY 296 LOGLDYFGPGVGLLGLVYSPVITVGLHOSPPTEL-ELFNQGSFIFATASMANIAQG 354
Db 278 YFTINLSPVAGAPWGAFAVQVILVIFGHWGVEPTMINLSRYGRTMIAMVGPSPFAQA 337
QY 355 AACLAVFFLAKSEKILKLAGAGSVAVLGITEPAIFGNLRLRWPFYIGIGTAAIGGALI 414
Db 338 GASLGVFLTKKPEVKAIAGSAALGFFGITEPSIYGVTLKYKKPFVIASIAAGGAIV 397
QY 415 ALFDIKAVAGLGAAGFLGVVVSIDAPDV---MFLYCAVVFTEVAFGAATAY-----GLYIV 466
Db 398 -----GAAGSSGAANA-IPGLILPFIIGKGVGFIL--GIAYAILSAIGTYFF 444
QY 467 RRGSDIDDATAAPVPAGTTKAEAPAEFSDSTIIQAPLGEAIALSSVSDAMFASGK 526
Db 445 ---GYKDEADGI---APTTEAKETGVE---AEVIVSPIRGNVPLNEVKDEAFSAGL 494
QY 527 LSGVAIVTKQLVSPVSGKIIVAPPSGHAFVTRKAEKDSNVDILMHIGDVTNLANGT 586
Db 495 LKGVAIVPQEGKLISPVNGTITETAPTGHAGIRS-----DKGVEILLHVGFDTVOLNGK 550
QY 587 HENPLKKQDEVKAGBELLCEFDIDAIAKAGYEVTTPIVVSN 627
Db 551 YFKLLVAQGRVLVVGQALFEFLDEIAKADGYDITTPIVVTN 591

RESULT 13
E98067
phosphotransferase system enzyme II (EC 2.7.1.69) scra [imported] - Streptococcus pneumo
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98067
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mahren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98067
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00370.1; PID:g15459232; GSPDB:GN00174
C:Genetics:
A:Gene: scra
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase
C:Keywords: phosphotransferase

Query Match 28.6%; Score 956; DB 2; Length 627;
Best Local Similarity 33.3%; Pred. No. 3.6e-57;
Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;
QY 1 MDHXDLAQRILRDGGEDNIVAAACATRLRLVLKDKVDVDRQSLDDPDLKGTFTFGGM 60
Db 1 MNQEIARKVLDAGGRENNSVAHCATRLRVVVVVVDEKINKEVILEKVVQGAFFNSGQ 60
QY 61 PQITVGPQDVHVFKE---LDDATSKDIAVSTEQLKQDVANNANWFSRAVKVLADIAPVL 117

Db 61 YQIFGTGTNKMVDEYVVLGLPTS-----SKDMKAEVAQGNWFQRAIRTEGDFVFP 115
QY 118 IPILVGGGLMANNVLAODLFGPOSIVEMFQISGVAEMINLMASAPFAFLPVLVG 177
Db 116 IPVIVATGLFWGVGFNALEM-----PLPGDFATYITQILTDTAIFLLPGLVWS 165
QY 178 ATRKFGNEFLGAGIGMAVFPVLVNGYDVAATAGEMPMSLFLGLDVAQAGYQGTVLP 237
Db 166 TFRVFGNPAVGIVGLMMLVSGSLPNAWAVA---QQGEVTAMNFFGP-IPVGLQGSVLP 221
QY 238 VLVVSWILATIEKFLHKRLMGTDFLITPVLTLTLGFTLFTIAIGPAMRWGDLIAHG 297
Db 222 AFILGVVAKFEKAVRKVPDVIDLVPFVTLVNSILGLFVIGPVFHVVENVILIATK 281
QY 298 GLYDFGPGVGLLGLVYSPVITVGLHOSPPTEL-ELFNQGSFIFATASMANIAQGAA 356
Db 282 AILSMPLGLGFLGGVHQVLIWVSGVHHIFNLLEVQLLAADHANPFAITTAATAAGAA 341
QY 357 CLAYFFLAKSEKILKLAGAGSVAVLGITEPAIFGNLRLRWPFYIGIGTAAIGGALIAL 416
Db 342 TVAVGVTKNPKLTLAPPAALSFLGITEPAIFGNLRLRWPFYIGIGTAAIGGGLASI 401
QY 417 FDIKAVAGLGAAGFLGVVVSIDAPDV---MFLYCAVVFTEVAFGAATAYGLYIVLR 468
Db 402 -----LGLAGTNGTITIPGTMLYVNGQLPOYLLMVAVSFALGALTMYMEFY----- 449
QY 469 NGSDIDDATAAPVPAGTTKAEAPAEFSDSTIIQAPLGEAIALSSVSDAMFASGKL 527
Db 450 --EVEDDATAAKAEVAEKEEVAAPALQNETIV--TPIVGDVVVALADNDVPFSSGAM 505
QY 528 GSGVAIVTKQLVSPVSGKIIVAPPSGHAFVTRKAEKDSNVDILMHIGDVTNLANGT 587
Db 506 GGGIADKESQGVVVALADAEVSIAPPTGHAFGLKTR-----NGAEVLIVHIGDVTSMNGDG 561
QY 588 FNPLKKQDEVKAGBELLCEFDIDAIAKAGYEVTTPIVVSN---YKKTGPNVNTYGLGIEA 644
Db 562 FEAKVAQGNKVRKAGDVLGTGFDNSKNKIAAAGLDDTTMVIIVTADVASVAPVAT---GVSVK 618
QY 645 GANLNV 651
Db 619 GDAVIEV 625

RESULT 14
C5220
trehalose PTS system, IIABC components [imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C5220
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C5220
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75956.1; PID:g14973388; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI884
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransfera
Query Match 28.4%; Score 949; DB 2; Length 655;
Best Local Similarity 34.9%; Pred. No. 1.1e-56;
Matches 234; Conservative 128; Mismatches 253; Indels 56; Gaps 19;
QY 7 AQRILRDIGEDNIVAAACATRLRLVLKDKVDVDRQSLDDPDLKGTFTFGMFOIIVG 66
Db 8 AKDLQAGGKENVATVATHTCATRMREFVLDGDDKANKVKAIESIPAVKGTFTNAGQFQVILG 67

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 01:26:49 ; Search time 24 Seconds
(without alignments)
1295.194 Million cell updates/sec

Title: US-10-019-284A-2
Perfect score: 3342
Sequence: 1 MDHKDLAQRLTRDIGGEDNI.....IEAGANLLNVAKKEAVPATP 661

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1609.5	48.2	651	PTSA_PEDPE	P43470 pediococcus
2	1478.5	44.2	664	PTSA_STRMU	P12655 streptococ
3	1277.5	36.7	480	PTSB_STAXY	P51184 staphylococ
4	1100.3	32.9	474	PTSB_PASMU	Q8c122 pasteurilla
5	1069.3	32.0	479	PTSB_VIBAL	P22825 vibrio algi
6	999.3	29.9	470	PTSB_BACSU	P39794 bacillus su
7	901.3	27.0	609	PTBA_BACSU	P40739 bacillus su
8	896.3	26.8	631	PTBA_ERWCH	P26207 erwinia chr
9	824.3	24.7	625	PTBA_ECOLI	P08722 escherichia
10	819.3	24.5	456	PTSB_SALTY	P08470 salmonella
11	816.3	24.4	456	PTSB_KLEPN	P27219 klebsiella
12	751.5	22.5	460	PTSB_BACSU	P05306 bacillus su
13	725.5	21.7	683	PTNA_CORGL	Q46072 corynebacte
14	712.5	21.3	674	PTGA_CORGL	Q45298 corynebacte
15	706.3	21.1	459	SACX_BACSU	P15400 bacillus su
16	655.3	19.6	473	PTBA_ECOLI	P36672 escherichia
17	438.3	13.1	474	YFEV_ECOLI	P77272 escherichia
18	437.3	13.1	482	PTSB_VIBCH	Q9kvd9 vibrio chol
19	435.5	13.1	485	PTDA_ECOLI	P24241 escherichia
20	353.3	10.6	699	PTGA_BACSU	P20166 bacillus su
21	329.3	9.8	651	PTAA_KLEPN	P45604 klebsiella
22	314.5	9.4	726	PTGA_STRPN	P35595 streptococ
23	313.5	9.4	631	YFVS_BACSU	P39816 bacillus su
24	308.5	9.2	648	PTAA_ECOLI	P09323 escherichia
25	299.5	9.0	324	PTGA_BACST	P42015 bacillus st
26	286.3	8.6	634	LACY_STRTR	P23936 streptococ
27	256.3	7.7	189	PTGA_BORBU	Q48480 borrelia bu
28	255.3	7.6	168	YFOE_BACSU	P30829 bacillus su
29	255.3	7.6	627	LACY_LACDE	P22733 lactobacill
30	249.5	7.5	154	PTGA_MYCCA	P45618 mycoplasma
31	246.3	7.4	168	PTGA_ECOLI	P08837 escherichia
32	244.3	7.3	168	PTGA_SALTY	P02908 salmonella
33	240.5	7.2	641	RAPP_PEDPE	P43466 pediococcus

ALIGNMENTS

RESULT 1

ID	PTSA_PEDPE	STANDARD;	PRT;	651 AA.
AC	P43470;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	PTS system, sucrose-specific IIBC component (EIIABC-SCR) (Sucrose-			
DE	permease IIBC component) (Phosphotransferase enzyme II, ABC			
DE	component) (EC 2.7.1.69) (EII-SCR).			
GN	SCRA.			
OS	Pediococcus pentosaceus.			
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae; Pediococcus.			
OX	NCBI_TaxID:1255;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PP1.0;			
RA	Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;			
RL	Submitted (XX-1994) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT			
CC	SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE			
CC	-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE			
CC	AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY			
CC	PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS			
CC	PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO			
CC	THE SUGAR.			
CC	-1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein			
CC	histidine + sugar phosphate.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- SIMILARITY: Contains 1 PTS EIIA domain.			
CC	-1- SIMILARITY: Contains 1 PTS EIIB domain.			
CC	-1- SIMILARITY: Contains 1 PTS EIIC domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z32771; CAA83668.1; -			
DR	EMBL; L32093; AAA25567.1; -			
DR	PIR; S44257; S44257.			
DR	HSP; P20166; IGPR.			
DR	InterPro; IPR001127; PTS_EIIA.			
DR	InterPro; IPR001996; PTS_EIIB.			
DR	InterPro; IPR003352; PTS_EIIC.			
DR	Pfam; PF00358; PTS_EIIA_1.			
DR	Pfam; PF00367; PTS_EIIB_1.			
DR	Pfam; PF02378; PTS_EIIC_1.			
DR	ProDom; PD002243; PTS_EIIA; 1.			
DR	ProDom; PD001476; PTS_EIIB; 1.			
DR	TIGRFAMS; TIGR00826; EIIB_glc; 1.			
DR	TIGRFAMS; TIGR00830; PTBA; 1.			
DR	PROSITE; PS00371; PTS_EIIA_1; 1.			

RESULT 2

KW Phosphorylation; Transmembrane; Complete proteome.
 FT DOMAIN 1 40 EIIB DOMAIN.
 FT DOMAIN ? ? EIIB DOMAIN.
 FT DOMAIN 533 664 EIIC DOMAIN.
 FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 331 331 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 585 585 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 664 AA; 69988 MW; 809P63E32281A9A1 CRC64;

 Query Match 44.2%; Score 1478.5; DB 1; Length 664;
 Best Local Similarity 46.0%; Pred. No. 1.9e-82;
 Matches 311; Conservative 127; Mismatches 203; Indels 35; Gaps 13;

 QY 1 MDHKLQAILRDIGEDNIVAAAHCAATRLRLVLKDTKDVRQSDLDLPGTETGGM 60
 Db 1 MDXSVASEVITAV-GKDNLVAAAHCAATRLRLVLKDDSKVDQKALDKNADVAGTKTDQ 59
 QY 61 FQIIVPGDGDVHVKELDDATSKDIADVSTEQKLDVVANNA--NWFSRAVKVLAIFVPLI 118
 Db 60 YQVIIGPGDGVNFYDEIRIKQTGL-TEVSTDLLKTAASGKKFNPMALIKLLSDIFVPII 118
 QY 119 PILVGGGLLMAINNVVAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 178
 Db 119 PALVAGGLMALNNFTLSGLSGTKSLVQOFPPIKGGSDMIOLMSAAPFWFLPILVISA 178
 QY 179 TKRFGNEFLGAGIMANVFPILVNGDYVAA-----TWTAGEMP-MWSEFLGDLVAQAGY 231
 Db 179 AKRFGANQFLGASIGIMVAPGAANTIGLAANAPISKAATIGAYTGFNIFGLHVTQASY 238
 QY 232 QGVNLPVLYVSWTALRFLKRLMGTAFLITPTVLTLLGFLTFIATGAPAMRWGDL 291
 Db 239 TYQVIVPVAWMLLSILEFFHKRLPSAVDFTFPLLSVIITGELTFIIVGPMKEVSDM 298
 QY 292 LAHGLQGLDFGPGVGLLFLGLVYSPVITGLHQSPPILEL---FNQG---GSPFAT 345
 Db 299 LTNGIYWLVDTTGFLGMVFGALYSPVMTGLHQSPPAIETQLISAFQNGTGHGDFIFT 358
 QY 346 ASMANAQAACLAFLVFLAKSEKLGKLAGASGVSAVLGITEPAIFGVNLRRLRWPFFVIG 405
 Db 359 ASMANVAQAATTAIFLFLKDKKMLGSSGVSALLGITEPAIFGVNLRKYRPFECALI 418
 QY 406 TAIGGALIALFDIKVAALGAAGFLGWSIDAPDMVFLVCVAVFVIAFGAAIAGLYL 465
 Db 419 GSASAAIAQLLOWAVSLGSAFLGLFTKASSIPFYVCELISPAIAFAVTVYGV--- 475
 QY 466 VRRNGSIDDPATAAPVPAAGTTKAEAPAEFSDST-----LIQAPLTGEAIALSSVS 518
 Db 476 --KTKAVDVFAEAAREAEAEVQ-EIPEEAASANKAQTDEVLAPLAGAEVETSYN 532
 QY 519 DAMFASGLGSGVAIVPTKQLVSPVSGKIVVAFPSGHAFAYKTAEDGNSVDILMHIGF 578
 Db 533 DPVFSEAMGKGAIKPSGNTVYAPVDGTVQIAFDTHAYGI--KSDNGA--EILTHIGI 588
 QY 579 DTVNLNGTHFNPLKQGDVEYKAGELLCEFDIDAIAKAGYEVTTPVIVSNYKKTGPVNTY- 637
 Db 589 DTVSMEGKGFQKQVQADQKIKKGDVLGTFDSKIAEAGLNTWTFIVTADYASVETLA 648
 QY 638 GLGEIEAGANLLNVAK 653
 Db 649 SSGTVAVGDSILEVKK 664

 RESULT 3
 PTSEB_STAXY
 ID PTSEB_STAXY STANDARD; PRT; 480 AA.
 AC P51184;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PTS system, sucrose-specific IIBC component (EIIB-SCR) (Sucrose-
 permease IIBC component) (Phosphotransferase enzyme II, BC component)
 DE (EC 2.7.1.69) (EIIB-SCR).
 GN SCRA.

OS Staphylococcus xylosus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1288;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=DSM 20267 / Isolate C2A;
 RA MEDLIN=94049686; PubMed=8232209;
 RX Wagner E., Goetz F., Brueckner R.;
 RT "Cloning and characterization of the scrA gene encoding the sucrose-
 specific Enzyme II of the phosphotransferase system from
 RT Staphylococcus xylosus";
 RL Mol. Gen. Genet. 241:33-41(1993).
 CC !- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.
 CC !- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
 CC histidine + sugar phosphate.
 CC !- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC !- SIMILARITY: Contains 1 PTS EIIB domain.
 CC !- SIMILARITY: Contains 1 PTS EIIC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X69800; CA49461.1; -
 DR PIR: S39978; S39978.
 DR InterPro: IPR001996; PTS_EIIB.
 DR InterPro: IPR003352; PTS_EIIC.
 DR Pfam: PF00367; PTS_EIIB; 1.
 DR Pfam: PF02378; PTS_EIIC; 1.
 DR ProDom: PD001476; PTS_EIIB; 1.
 DR ProDom: PD001476; PTS_EIIB; 1.
 DR TIGRfam: TIGR00826; EIIB-9lc; 1.
 DR PROSITE: PS01035; PTS_EIIB-CYS; 1.
 DR Phosphotransferase system; Sugar transport; Transferase;
 KW Transmembrane; Inner membrane; Phosphorylation.
 FT DOMAIN 1 ? EIIB DOMAIN.
 FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 325 325 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 480 AA; 51326 MW; AB4E1D9765D84E47 CRC64;

 Query Match 36.7%; Score 1227; DB 1; Length 480;
 Best Local Similarity 52.7%; Pred. No. 2.3e-67;
 Matches 252; Conservative 83; Mismatches 133; Indels 10; Gaps 6;

 QY 1 MDHKLQAILRDIGEDNIVAAAHCAATRLRLVLKDTKDVRQSDLDLPGTETGGM 60
 Db 1 MDXSVASEVITAV-GKDNLVAAAHCAATRLRLVLKDDSKVDQKALDKNADVAGTKTDQ 59
 QY 61 FQIIVPGDGDVHVKELDDATSKDIADVSTEQKLDVVANNA--NWFSRAVKVLAIFVPLIPI 120
 Db 61 YQVIIGPGDGVNFYDEIRIKQTGL-TEVSTDLLKTAASGKKFNPMALIKLLSDIFVPII 119
 QY 121 LVGGGLLMAINNVVAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 179
 Db 120 IVAGGLLMAINNVVAQDLFGPQSLVEMFQIISGVAEMINLMASAPFAFLPVLGFTA 179
 QY 180 KRFGGNEFLGAGIMANVFPILVNGDYVAA--EMPMWSLFLGDLVAQAGYQGTETGGM 238
 Db 180 KRFGGNAYLGAALGMLVHPPELMSAYDYPKALEAGKEIPHNLFGLINQVGYQGVLP 239
 QY 239 LVVSWILATIEKHLKRLMGTAFLITPTVLTLLGFLTFIATGAPAMRWGDLAHGQ 298
 Db 240 LVATVILATIEKHLKRLMGTAFLITPTVLTLLGFLTFIATGAPAMRWGDLAHGQ 299

QY 300 YFGGPGVGLGLGLYSPVITVIRGLHOSFPPIELELFN-OGGSFIPATASMANIAGAAAL 358
 Db 282 YQAVPAFAAGVGGFWQIFVMEFLHGLVGLVCINNFTVLGVDTPMLLPAALMAQVGAAL 341
 QY 359 AVFELAKSEKLGLAGAGSVAVLGTTPAIFGVNLRWRPPIYIGTAAIGGALIALFD 418
 Db 342 GVFLCERDAQKVVAGSAALTSFGITBPAYGVNLPKRPYVACISGALGATIGYAQ 401
 QY 419 IKAVALGAAG-FLGVVSDADPMVMFLVCANVTFVIAFGAAIAYGLYLVRNNGSIDPAT 477
 Db 402 TKVYSEGLPSITFTQIPSTGIDFTVWASVGGVIAIGCAPGVGVMLHFTAKRQP-AQ 460
 QY 478 AAPVPAGTAKAEAPAEBSNDSTIOAPLTGCEATLSSVSDAMFASGKLGSGVAIVPTK 537
 Db 461 GAP--QKTEPVITPPEGG---ICSPMTGEIVPLHVADTTFASGLLKGIALPSV 513
 QY 538 GOLVSPVSKVIVVPPSGHAFVAVRTKAEDGSNDVILMHIGFTVNLNGTHFNPLKKQDGE 597
 Db 514 GEVRSPVAGRIASLFTLAIGI--ESDDG--VEILIHVGIDTVKLDGKFFSARHNVGDK 569
 QY 598 VRAGELLCFEDDAIKAAQYEVVTPVIVSNYKKTGPVNTYGLGEIERGANLLVAK 653
 Db 570 VNTGDLISFDIPATREAGFDLTPVLIISDDFTDVLPHGTAQISAGEPLLSIIR 625

RESULT 10

PTSB_SALTY STANDARD; PRT; 456 AA.
 AC P08470;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
 DE permease IIBC component) (Phosphotransferase enzyme II, BC component)
 DE (SC 2.7.1.69) (EII-SCR).
 GN SCRA.
 OS Salmonella typhimurium.
 OG Plasmid pUR400.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88216186; PubMed=3285123;
 RA Ebner R., Lengeler J.W.;
 RT "DNA sequence of the gene scrA encoding the sucrose transport protein
 RT EnzymeII(Scr) of the phosphotransferase system from enteric bacteria:
 RT homology of the EnzymeII(Scr) and EnzymeII(Bgl) proteins.";
 RL Mol. Microbiol. 2:9-17(1988).
 RN [2]
 RP REVISIONS.
 RX MEDLINE=94018607; PubMed=8412665;
 RA Jahreis K., Lengeler J.W.;
 RT "Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid
 RT repressor for sucrose and D-fructose specific regulons from enteric
 RT bacteria.";
 RL Mol. Microbiol. 9:195-209(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=36186840; PubMed=8628219;
 RA Titgemeyer F., Jahreis K., Ebner R., Lengeler J.W.;
 RT "Molecular analysis of the scrA and scrB genes from Klebsiella
 RT pneumoniae and plasmid pUR400, which encode the sucrose transport
 RT protein Enzyme II Scr of the phosphotransferase system and a
 RT sucrose-6-phosphate invertase.";
 RL Mol. Gen. Genet. 250:197-206(1996).
 RN [4]
 RP SEQUENCE OF 1-7 FROM N.A.
 RX STRAIN=6153-62;
 RC MEDLINE=91100329; PubMed=1846143;
 RA Hardesty C., Ferran C., Dizenzo J.M.;
 RT "Plasmid-mediated sucrose metabolism in Escherichia coli:
 RT characterization of scrY, the structural gene for a

phosphoenolpyruvate-dependent sucrose phosphotransferase system
 outer membrane porin.";
 RL J. Bacteriol. 173:449-456(1991).
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS). A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HEP). IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR. EIIIC-SCR BELONGS TO THE EIIIC DOMAIN TYPE AND LACKS A
 CC COVALENTLY BOUND EIIA DOMAIN. INSTEAD, EII-SCR-MEDIATED
 CC PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
 CC GLC. A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
 CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
 CC histidine + sugar phosphate.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -!- SIMILARITY: Contains 1 PTS EIIIC domain.
 CC -!- SIMILARITY: Contains 1 PTS EIIIC domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Y00541; CRA68605.1; ALT_SEQ.
 DR EMBL; X57750; CRA47973.1;
 DR EMBL; M38416; AAA98418.1;
 DR PIR; S01036; WQEBST.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR InterPro; IPR004719; PTSIIC_glc.
 DR Pfam; PF00367; PTS_EIIB; 1.
 DR Pfam; PF02378; PTS_EIIC; 1.
 DR Prodom; PD001476; PTS_EIIB; 1.
 DR TIGRFAMs; TIGR00826; EIIIC_glc; 1.
 DR TIGRFAMs; TIGR00852; pts-Glc; 1.
 DR PROSITE; PS01035; PTS_EIIB-CYS; 1.
 DR KW Phosphotransferase system; Transferase; Transport; Sugar transport;
 KW Plasmid; Transmembrane; Inner membrane; Phosphorylation.
 FT DOMAIN 1 111
 FT DOMAIN 112 456
 FT TRANSMEM 112 132
 FT TRANSMEM 144 164
 FT TRANSMEM 181 201
 FT TRANSMEM 213 233
 FT TRANSMEM 247 267
 FT TRANSMEM 288 308
 FT TRANSMEM 330 350
 FT TRANSMEM 360 380
 FT TRANSMEM 388 408
 FT TRANSMEM 428 448
 FT MOD_RES 26 26
 FT MOD_RES 309 309
 FT SEQUENCE 456 AA; 47854 MW; 98A6F1620AE50885 CRC64;

Query Match 24.5%; Score 819; DB 1; Length 456;
 Best Local Similarity 38.2%; Pred. No. 1.le-42;
 Matches 180; Conservative 84; Mismatches 179; Indels 28; Gaps 6;
 QY 1 MDKDLAQRILRDIGGEDNIVAAAHCAHRLRLVLKDTKDVDRQSLDDPDLDKGFETGGM 60
 Db 1 MDEFQISCSLLPLLGKKNIAAASAAHCAHRLRLVLVDDSLADQQAIGKGVKGGCFRNAG 60
 QY 61 FOITVGPQGDVHVFKELDDATSKDIAVSTEQCLKDVANNWFSRAVKLADIFVPLPI 120
 Db 61 MQIIFGTGVNVKVAFTQAAGISES-SKSEAADTAARKLAPFQIARLLNIFVPIIPA 119
 QY 121 LVGGGLLMAINNVLVAQDLFGPQSLVEMFPQISGVAEINLMASAPFAFLPVLGFTATK 180
 Db 120 IVASGLLGLLGMVKTYGWVDPGNAIYI-----MLDMCSSAAFIILPILIGFTAAR 170

Qy 181 RFGNEFLGAGIGAMVFTLVNGYDVA--TMTAGEMPMSLFGLDVAQAGYQGTVP 237
 Db 171 EFGNGPYLGATLGGLIHTHPALTANAGVAGFHTM-----NFFGEIAMIYQGTVP 222
 Qy 238 VLVVSWILATIEFLHKLRLMGTADELITPVLTLTLTGLFTFTIAIGPAMRWGDLAHLGQ 297
 Db 223 VLLAVWFMISVEKQLRAIPDALDLTLTFTLVIISSFIALLIIPAGRALGDSIFVLS 282
 Qy 298 GLYDFGPGVGLLFLGLVYSPVITGLHQSPFPPIELELFNQ---GGSFIATASMANIAOG 354
 Db 283 TLISHAGWLAGLFLGLYVIVITGHSFHAIVEAGLLGNPSIGVNFLLPIWAMANVAG 342
 Qy 355 AACLAVFFLAKSEKLKLAGAGSVAVLGITTEPAIFGVNLRWRPFIYIGITAAIGGALI 414
 Db 343 GACFVFWKTKDKIKATILPSAFSAMLGITEAIFGINLRFVKPIALVGAAGAV 402
 Qy 415 ALFDIKAVLGAAGFLGVSDIAPDMVMFVCAVTFVTAFAAIAIYL 465
 Db 403 VSVHVTAVGLTAIPGMATVQASSLLNY----IIGWIAFGVFTVSLVL 449

RESULT 11
 PTSE_KLEPN
 ID PTSE_KLEPN STANDARD; PRT; 456 AA.
 AC P27219;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
 DE permease IIBC component) (Phosphotransferase enzyme II, BC component)
 DE (EC 2.7.1.69) (EII-SCR).
 GN SCRA.

OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_taxid=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1033-5p14 / KAY2026;
 RX MEDLINE=9618840; PubMed=8628219;
 RA Tigemeyer F., Jahreis K., Ebner R., Lengeler J.W.;
 RT "Molecular analysis of the scrA and scrB genes from Klebsiella
 pneumoniae and plasmid pUR400, which encode the sucrose transport
 RT protein Enzyme II scr of the phosphotransferase system and a
 RT sucrose-6-phosphate invertase".
 RL Mol. Gen. Genet. 250:197-206(1996).

CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOTOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A
 CC COVALENTLY BOUND EIIB DOMAIN. INSTEAD, EII-SCR-MEDIATED
 CC PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-
 CC GIC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.
 CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
 CC histidine + sugar phosphate.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -!- SIMILARITY: Contains 1 PTS EIIB domain.
 CC -!- SIMILARITY: Contains 1 PTS EIIC domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X57401; CAA40658.1; .
 DR PIR: S62331; S62331.

DR InterPro: IPR001396; PTS_EIIB.
 DR InterPro: IPR003352; PTS_EIIC.
 DR InterPro: IPR004719; PTS_EIIC_glc.
 DR Pfam: PF00367; PTS_EIIB; 1.
 DR Pfam: PF02378; PTS_EIIC; 1.
 DR TIGRFAMS: TIGR00826; EIIB_glc; 1.
 DR TIGRFAMS: TIGR00852; pts-glc; 1.
 DR PROSITE: PS01035; PTS_EIIB-CYS; 1.
 KW Phosphotransferase system; Transport; Sugar transport; Transferase;
 KW Transmembrane; Inner membrane; Phosphorylation.
 FT DOMAIN 1 111 EIIB DOMAIN.
 FT DOMAIN 112 456 EIIC DOMAIN.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 247 267 POTENTIAL.
 FT TRANSMEM 288 308 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 FT TRANSMEM 360 380 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 FT TRANSMEM 428 448 POTENTIAL.
 FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 456 AA; 48022 MW; 4AFDF3405CAEFC66 CRC64;

Query Match 24.4%; Score 816; DB 1; Length 456;
 Best Local Similarity 37.8%; Pred. No. 1.7e-42;
 Matches 176; Conservative 88; Mismatches 177; Indels 24; Gaps 5;

Qy 1 MDKOLAQRILRDIGGEDNIVAAAHCAHRLRLVLDKTDVDRQSLDDDDPLKGFETGGM 60
 Db 1 MDEQISRLSLPLLGKENIASAAHCAHRLRLVLDKTDVDRQSLDDDDPLKGFETGGM 60
 Qy 61 FOIIVGPGVDVHVKELDDATSKDIATVSTEQLKOVANNWFSRAVKVLADIFVPLPI 120
 Db 61 MQIIFGTGVNVKYAFAFIOQAAGISE-SKSEAADAAKKNLPFQRIARLLSNIFPIIPA 119
 Qy 121 LVGGGLLMAINVLVAQDLFGPQSVLEMPPOISQSVAEINMLASAPFAFLPVLVGFATK 180
 Db 120 IVASGLLGLLGNVKTGYGVDPNSALYI-----MLDMCSSAAFIILPILIGFTAAR 170
 Qy 181 RFGNEFLGAGIGAMVFTLVNGYDVA--TMTAGEMPMSLFGLDVAQAGYQGTVP 237
 Db 171 EFGNGPYLGATLGGLIHTHPALTANAGVAGFHTM-----NFFGEIAMIYQGTVP 222
 Qy 238 VLVVSWILATIEFLHKLRLMGTADELITPVLTLTLTGLFTFTIAIGPAMRWGDLAHLGQ 297
 Db 223 VLLAVWFMISVEKQLRAIPDALDLTLTFTLVIISSFIALLIIPAGRALGDSIFVLS 282
 Qy 298 GLYDFGPGVGLLFLGLVYSPVITGLHQSPFPPIELELFNQ---GGSFIATASMANIAOG 354
 Db 283 TLISHAGWLAGLFLGLYVIVITGHSFHAIVEAGLLGNPSIGVNFLLPIWAMANVAG 342
 Qy 355 AACLAVFFLAKSEKLKLAGAGSVAVLGITTEPAIFGVNLRWRPFIYIGITAAIGGALI 414
 Db 343 GACFVFWKTKDKIKATILPSAFSAMLGITEAIFGINLRFVKPIALVGAAGAV 402
 Qy 415 ALFDIKAVLGAAGFLGVSDIAPDMVMFVCAVTFVTAFAAIAI 459
 Db 403 VSVHVTAVGLTAIPGMATVQASSLLNYIIGWIAFGVFTVSLVL 447

RESULT 12
 PTSE_BACSU
 ID PTSE_BACSU STANDARD; PRT; 460 AA.
 AC P05306;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
 DE permease IIBC component) (Phosphotransferase enzyme II, BC component)
 DE (EC 2.7.1.69) (EII-SCR).

GN SACP OR IPA-49D.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88097369; PubMed=3122206;
RA Fouet A., Arnaud M., Klier A., Rapoport G.;
RT "Bacillus subtilis sucrose-specific enzyme II of the
RT phosphotransferase system: expression in *Escherichia coli* and
RT homology to enzymes II from enteric bacteria";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8773-8777(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boursier L., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerison P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Guim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Hega K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Porcettelle D., Porcettelle S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rochelle B., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serrier P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Weitzenecker T.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Yasumoto K., Yata K.,
RA Winters P., Wibat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*";
RL Nature 350:249-256(1997).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -1- FUNCTION: ACTS AS BOTH A KINASE AND A PHOSPHATASE ON SACT (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Contains 1 PTS EIIB domain.
CC -1- SIMILARITY: Contains 1 PTS EIIC domain.
CC -1- SIMILARITY: 57% IDENTITY WITH THE NEGATIVE REGULATORY PROTEIN OF

SACP.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: J03006; AAA22727.1; -
EMBL: X73124; CAA51605.1; -
EMBL: Z99123; CAB15831.1; -
EMBL: A39338; A39338.
HSP: P05053; IIBA.
DR Subtilisin; BGI0595; sactP.
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR InterPro: IPR004719; PTSIIC_glc.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR TIGRFAMS: TIGR00826; PTS_EIIB; 1.
DR TIGRFAMS: TIGR00852; PTS_EIIC; 1.
DR ProSITE: PS01035; PTS_EIIB_CYS; 1.
DR Phosphotransferase system; Sugar transport; Transferase;
KW Transmembrane; Phosphorylation; Complete proteome.
FT DOMAIN 1 ? EIIB DOMAIN.
FT DOMAIN ? 460 EIIC DOMAIN.
FT MOD_RES 26 26 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 247 267 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 328 348 POTENTIAL.
FT TRANSMEM 358 378 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 429 449 POTENTIAL.
SQ SEQUENCE 460 AA; 49002 MW; 6A303DC042BFE379 CRC64;
Query Match 22.5%; Score 751.5; DB 1; Length 460;
Best Local Similarity 38.2%; Pred. No. 1.4e-38;
Matches 178; Conservative 86; Mismatches 175; Indels 27; Gaps 7;
QY 1 MDHKLQRIILRDIGGEDNIVAAAHCAIRLRVLKDTKDVRQSLDDDDPDLKGTFTGGM 60
DB 1 MDYKETAIRLLIGGGKENIISAHCAIRLRVLMKDESKIDQAOVELDGVKAFSSGQ 60
QY 61 FOIIVGPGVDHVKELDDATSKDIATVSTEQ---LKDVAANNWFSRAVKVLADIFVPL 117
DB 61 YQIIFGTGLVAKVF---DAFSKEADIEREHHVNHQDAKEKLPAAAFATLSNIEFPI 116
QY 118 IPIVGGGLLMAINNLVLAQDLFGPSLVEMFPQISGVAEMINLMSAFAPFLVLYGFT 177
DB 117 IPIVAGLLMG-----LLGMNFAHMSKDSALLQLDMFSSAAFIPLIGVS 167
QY 178 ATKRGGNEFIAGIGMAWVPTLVNGVDVAATVATAGEMPWVSLFGLDVAAGYQGVLP 237
DB 168 ASKEFGSPYLGAVTGGGINIHPLNPNWGLAE-----EQLITCFSDLSIFSGTNCYP 222
QY 238 VLWVSWILATTEKFLKRLMGTAFLTPVLTLTLTGFTETAIAGPAMRWGDLHLAQ 297
DB 223 CPACGVCDDEQREMDEKSGSTCGDLLVTPFVTVVTFVAVTAVTAVTAVTAVTAVT 282
QY 298 GLYDFGGVGGLLFGLVYPIVITGLHSGFPFIELELF-NOGGSFIFATASMANIAQGA 356
DB 283 YVYDHAGFVAGLIFGGTYSILVLTGVHSHFAIEAGLIADICKNYLLPWSMANVQOGGA 342
QY 357 CLAVFFLAKSEKRLKGLAGASGVSAVLGITTEPAIFGVNLRWLPFVIGITGAIGGALIAL 416

```

Db 343 GLAVFWAKAKTKETALPAASAFGLGIEPIFGVNLRYRPFIAAMIGGALGGAYVVF 402
QY 417 FDIKAVAGAGFLGVVSDAP-----DMVFLVCVVTFVIAFGAA 458
Db 403 THVAANAYGLTG-IPMIAAAPFGSGLNLIHYLIGMAIAASAFIAA 447

RESULT 13
PTNA_CORGL
ID PTNA_CORGL STANDARD; PRT; 683 AA.
AC Q45072;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PFS system, mannose-specific IIABC component (EIIABC-Man) (Mannose-
DE permease IIABC component) (Phosphotransferase enzyme II, ABC
DE component) (EC 2.7.1.69) (EII-Man/EIIC-Man).
GN PTSM OR CGL1360.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
X1 [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=94314161; PubMed=8039653;
RA Lee J.K., Sung M.H., Yoon K.H., Yu J.H., Oh T.K.;
RT "Nucleotide sequence of the gene encoding the Corynebacterium
RT glutamicum mannose enzyme II and analyses of the deduced protein
RT sequence.";
RL FEMS Microbiol. Lett. 119:137-145(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOTRANSFERASE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IIIC DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar -> protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Contains 1 PTS EIIA domain.
CC -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L18874; AAA53546.1; .
CC EMBL; AP005278; BAB98753.1; .
CC HSSP; P08837; 1GIC.
CC InterPro; IPR001127; PTS_EIIA.
CC InterPro; IPR001996; PTS_EIIB.
CC InterPro; IPR003352; PTS_EIIC.
CC Pfam; PF00358; PTS_EIIA_1; 1.
CC Pfam; PF00367; PTS_EIIB; 1.
CC Pfam; PF02378; PTS_EIIC; 1.
CC ProDom; PD002243; PTS_EIIA; 1.
CC ProDom; PD001476; PTS_EIIB; 1.
CC TIGRFAMs; TIGR00830; PTBA; 1.
CC PROSITE; PS00371; PTS_EIIA_1; 1.
CC PROSITE; PS01035; PTS_EIIB_CYS; 1.

```

```

KW Phosphotransferase system: Sugar transport; Transferase;
KW Phosphorylation; Transmembrane; Complete proteome.
FT DOMAIN 1 43 EIIA DOMAIN.
FT ? 683 EIIA DOMAIN.
FT DOMAIN 126 146 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 376 396 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
FT MOD_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 602 602 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 683 AA; 72571 MW; 0B42CAEC60828075 CRC64;

Query Match 21.7%; Score 725.5; DB 1; Length 683;
Best Local Similarity 30.9%; Pred. No. Be-37;
Matches 216; Conservative 119; Mismatches 286; Indels 79; Gaps 18;

QY 7 AGRILRDIGGEDNIVAAAHCAATRLRLVLKDKDVRQSLDDDDPDLKGTFTG--GMFOII 64
Db 9 SCHILELUGDPNITSMTHCAATRLRPQVQKDSIVDQOEIDSPSVLGVVPGQSTGM-QVV 67
QY 65 VPGDGVHVFKEI-----DOATSKDIADVSTEQKQDVVANNANWFSRAVKVLADI 113
Db 68 MG-GSVANYVYQILKLDGKMKHFADEATES-----SSKKEYGVGRKYSWIDVAFELSDT 122
QY 114 FVPLPILVGGGLLMAINNVLAQDLFGPQSLVEMFQISGVAEMINLMASAPFAFLPVL 173
Db 123 FRPILWALLGSLIITL---LVLDATFGLDQDFRPMDEQPDITYVFLHSMWRSVFFLPIM 179
QY 174 VGFTATKRGGEFLGAGIGAMKVPFTLVNGVDVAATMTAGEMPMSFLGLDVAQGYOG 233
Db 180 VGATAARKLGANEWIGAAIPAAALLP-----EFLALGSAGDTV--TVFGLPWLNDYSG 231
QY 234 TVLPVLVSWILATIEKFLHKLMTADLIITPVLTLTLTGTLTIAIGPANKWGDLLA 293
Db 232 QVFPPLIAAGLYWVEKGLKPIEAVQMVVFPFSLIMIPATAFLGPPFGIGVNGIS 291
QY 294 HGLQGLYDFGGPVGGLLGLVYSPVITGLHQSFPPIELEFNQ--GGSFIFATASMANIA 352
Db 292 NLLEANNPSPFISIVIPLLYPLVPLGLHNPALNMIQNTLGYDFIQGPMGAWNPA 351
QY 353 QGAACLAVFELAKSEKLGLAGAS--GVSAYL--GITEPAIFGVNLRWRPFIIGITAA 408
Db 352 CFGLVTGVFLLSIKERNKAMQVSLGGLMAGLGGISEPSLYGLLRFKTKTYRLLPGCL 411
QY 409 IGGALIALFDIKAVAGAGFLGVVSDAPDMVFLVCVVTFVIAF-GAAIAYGLYLVR 467
Db 412 AGIYMGIFDIKAYAVFVLSLLTIPAMD-PWLYTIGIAYAVFVSNFLVLDY----R 465
QY 468 RNSID-----PDAT-AAVP-----PAGTKAKAEAPA 494
Db 466 SNEERDEARAKVAADKQAEEDLKAEANATPAAPVAAAGAGAGAGAGAAAGATAAKPK 525
QY 495 EFSNDSTIQAPLTGEAIALSVSDAMFASGKLGSCVAIVPTKQGLVSPVSKIVVPPS 554
Db 526 LAAGEVVDIVSPLEGRKAIPLEVPDIPFAAGKLGIPGIAOPTGNTVVAADTVILVQKS 585
QY 555 GHAFVAVTRAEQDSNVDILMHIGFTVNLNGTHFNPLKKQGDVEKAGELLCFFDIDA 614
Db 586 GHAVLRL---DSGVEILLVHVGDLTVQLGGEGFTVHVERQOVKAGDPLITFDADFIRS 641
QY 615 AGYEVTPIVVSNNYKKTGPVNTYGLGEIEAGANLLNVAKK 654
Db 642 KDLPLITPVVYSNAKFEIEGIPADQANSSTTVIKVNGK 681

RESULT 14

```


RA Steinmetz M.;
RT "Nucleotide sequence of the sacS locus of *Bacillus subtilis* reveals
RT the presence of two regulatory genes.";
RL Gene 90:153-155(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moser I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsch M., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Gusepki G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE OF 348-459 FROM N.A.
RC STRAIN=168;
RX MEDLINE=92216127; PubMed=1806041;
RA Glaser P., Kunst F., Debatbouille M., Vertes A., Danchin A.,
RA Dedonder R.;
RT "A gene encoding a tyrosine tRNA synthetase is located near *sacS* in
RT *Bacillus subtilis*.";
RL DNA Seq. 1:251-261(1991).
CC -!- FUNCTION: Negatively regulates *sacY*.
CC -!- SIMILARITY: Contains 1 PTS EIIB domain.
CC -!- SIMILARITY: Contains 1 PTS EIIC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29333; AAA75335.1; -;
DR EMBL; X52480; CAA36719.1; -;

DR EMBL; Z99123; CAB15867.1; -;
DR EMBL; X73124; CAA51570.1; -;
DR PIR; JU0293; JU0293.
DR HSSP; P05053; 1IBA.
DR Subtilist; BG10560; sacX.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR TIGRfam; TIGR00826; EITB_glc; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
KW Transcription regulation; Transferase; Phosphorylation;
KW Complete proteome.
FT DOMAIN 1 ? EIIB DOMAIN.
FT DOMAIN ? 459 EIIC DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 136 136 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 433 433 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 459 AA; 49024 MW; A5C4E996ECDA3D40 CRC64;

Query Match 21.1%; Score 706; DB 1; Length 459;
Best Local Similarity 35.9%; Pred. No. 7.9e-36;
Matches 168; Conservative 93; Mismatches 167; Indels 40; Gaps 11;

QY 3 HXDLAQRILRDIGGEDNIVAAAHCAATRLRLVLTQDKVDVRSQSLDDPDPLKGTFTGGMFQ 62
DB 2 HREIAKELLLAGGKNNIISISCHTTLRLDFDKDETAKIDHAIENLOGVGGFFRYGLFQ 61
QY 63 ITVPGGDVDFVKEL----DDATSKDIAYSTQLKDXVANNWFSSRAVKVLADIFVPLI 118
DB 62 IIFGAGVKNKIYKVHVHETAPSEE---PVHQKK--ASRKLNPAAAFKLTSDIFVPII 116
QY 119 PILVGGGILLMANNNVLVAQDLFGPQSLVEMFPQISGVAEMINLMASAPAFPLVVGFTA 178
DB 117 PAITASGLLGG-----LIGMIKVFHFWFAAGSPWIKMLDLVSTAFILLPLVGFSA 167
QY 179 TRFRGNEFLGAGIGMAMVFTLVNGDYVAATMTAGEMPMSLFLGLDVAQAGYQGTVLFPV 238
DB 168 ARQFGSNPYLGAIVAGLLTHPDLDP-SMLGSKTSSLDI--GLHIPMGYQGSMPIT 223
QY 239 LVVSVTLATIEKFLHKLMTGADFLTPVTLTLTLGFTFTFIAIGPAMRWVGGDLAHLQGG 298
DB 224 LLSVFMVSKIEKLLKSIVPKSLDVLPITVITVWVGCLALIVNPAASIIQIMTQSIYV 283
QY 299 LYDFGPGVGGLLFGLVYSPIVITGLHQSPPIELF---NQGGSFIFATASMANIAQGA 355
DB 284 IYDHAGIAAGALFGGIYIIVLSGLHHSFYATEATLLANPHVGVNPLVPIWMSMANVAQGG 343
QY 356 ACLAVFELAKSEKLGAGASGVSAVLGITEPAIEGVNLRRLRWPYIGIGTA---AIGA 412
DB 344 AGLAVFLKTKQSSLLKIALPASLTAFLGIVPIVGVNKLIRPF---IGAAIGGAIGA 400
QY 413 LIALFDIKAVCALGAGFLGVVSDADP-----DMVMFLVCAVVTFFV 452
DB 401 YVAVGVVANSYGLTG-IPMISIVLPFGAANFVHYMIGELIAAVSAFI 447

Search completed: October 1, 2003, 05:02:27
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 01:21:54 ; Search time 61 Seconds
(without alignments)
1719.972 Million cell updates/sec

Title: US-10-019-284A-2
Perfect score: 3342
Sequence: 1 MDHKLQAQLRLDIGEDNI.....IEAGANLLNVAKEVPATP 661

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3342	100.0	661	22	AAB69080
2	3310	99.0	661	22	Brevibacterium lac
3	2332	69.8	468	22	C glutamicum prote
4	2332	69.8	468	23	C. glutamicum phosph
5	1806	54.0	362	22	C. glutamicum meta
6	957	28.6	617	23	C. glutamicum phosph
7	956	28.6	627	22	ABBA47495
8	949	28.4	655	24	AAU37874
9	943	28.4	655	24	ABP81458
					S. pneumoniae type

10	935.5	28.0	620	23	ABP27215
11	931	27.9	676	23	Streptococcus poly
12	924.5	27.7	634	23	ABP25654
13	920	27.5	674	23	Listeria monocytog
14	880.5	26.3	639	23	Streptococcus poly
15	859	25.7	617	23	ABP27215
16	829	24.8	618	23	Listeria monocytog
17	822.5	24.6	636	23	ABBA49923
18	799.5	23.9	612	24	Lactococcus lactis
19	799.5	23.9	612	24	Streptococcus pneu
20	791	23.7	622	23	S. pneumoniae type
21	785.5	23.5	620	23	Streptococcus poly
22	747.5	22.4	640	23	Streptococcus poly
23	725.5	21.7	683	22	Listeria monocytog
24	725.5	21.7	683	22	C glutamicum prote
25	725.5	21.7	683	22	C. glutamicum phosph
26	719	21.5	494	23	C. glutamicum meta
27	691	20.7	704	22	Listeria monocytog
28	584	17.5	381	24	Propionibacterium
29	558.5	16.7	791	23	Streptococcus pneu
30	550	16.5	484	22	Bifidobacterium lo
31	538	16.1	455	22	Enterococcus faeca
32	529	15.8	334	24	Salmonella typhi c
33	528.5	15.8	243	23	S. pneumoniae type
34	522	15.6	484	22	Staphylococcus epi
35	517	15.5	484	22	Staphylococcus aur
36	511	15.3	521	23	Pathogen specific
37	511	15.3	478	23	Staphylococcus epi
38	490.5	14.7	439	22	Lactococcus lactis
39	477	14.3	461	23	Staphylococcus aur
40	440.5	13.2	454	23	Streptococcus poly
41	438	13.1	474	22	Lactococcus lactis
42	422.5	12.6	413	22	E. coli cellular p
43	404	12.1	523	22	C. glutamicum phosph
44	374	11.2	249	24	Propionibacterium
45	374	11.2	249	24	Streptococcus pneu
					S. pneumoniae type

ALIGNMENTS

RESULT 1
AAB69080
ID AAB69080 standard; Protein; 661 AA.
XX AAB69080;
AC AC
XX
DT 20-APR-2001 (first entry)
XX
DE Brevibacterium lactofermentum sucrose PTS enzyme II protein SEQ ID NO:2.
XX
KW Brevibacterium lactofermentum; sucrose PTS enzyme II; sucrose;
KW phosphoenolpyruvate:carboxylate phosphotransferase system; glucose;
KW coryneform bacterium; phosphoenolpyruvate-sugar transport system.
XX
OS Brevibacterium lactofermentum.
XX
XX
PN WO200102584-A1.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-JP04348.
XX
PR 02-JUL-1999; 99JP-0189512.
XX
FA (AJIN) AJINOMOTO CO INC.
XX
PI Izui M, Sugimoto M, Nakamatsu T, Kurahashi O;
XX
DR WPI; 2001-138150/14.
XX
XX N-PSDB; AAF32543.
XX
PI Gene encoding sucrose phosphoenolpyruvate-sugar transport system enzyme

PT II obtained by cassette ligation-mediated amplification of downstream
 PT domain of coryneform bacterium sucrose gene, with sucrose-binding
 PT activity -
 XX
 XX
 PS Claim 1; Page 29-32; 45pp; Japanese.
 XX
 CC The present sequence represents the Brevibacterium lactofermentum
 CC sucrose PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system
 CC or phosphoenolpyruvate-sugar transport system) enzyme II, which has
 CC sucrose-binding activity. A coryneform bacteria produced with the
 CC sucrose-PTS enzyme II gene can have more efficient sugar uptake and
 CC improved amino-acid and nucleic acid productivity. The sucrose PTS gene
 CC and it's disrupted gene, such as one without the sucrose PTS function,
 CC can be used to produce new breeds of coryneform bacterial strains to
 CC uptake sugar more efficiently e.g. glucose only or and sucrose, and can
 CC have improved amino-acid and nucleic acid productivity.
 XX
 XX Sequence 661 AA;

Query Match 100.0%; Score 3342; DB 22; Length 661;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 661; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MDHKLQAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60
 DB 1 MDHKLQAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60
 QY 61 FOIIVGPGDVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120
 DB 61 FOIIVGPGDVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120
 QY 121 LVGGGLLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVPGFTATK 180
 DB 121 LVGGGLLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVPGFTATK 180
 QY 181 RFGCNFLGAGIGMAMVFTLVNGYDVAATMAGEMPMSLFGDVAQAGYOGTVLPVLY 240
 DB 181 RFGCNFLGAGIGMAMVFTLVNGYDVAATMAGEMPMSLFGDVAQAGYOGTVLPVLY 240
 QY 241 VSVILATIEKFLHKLGMGTADFLITPVILLTGTFTTAIGPAMRWGDLAHGLQGLY 300
 DB 241 VSVILATIEKFLHKLGMGTADFLITPVILLTGTFTTAIGPAMRWGDLAHGLQGLY 300
 QY 301 DFGPGVGLLFGLVYSPVITGLHQSPFPIELFNQGSFIFATASMANIAQGAACLA 360
 DB 301 DFGPGVGLLFGLVYSPVITGLHQSPFPIELFNQGSFIFATASMANIAQGAACLA 360
 QY 361 FFLAKSEKLGAGASGVSAVLGITPAIFGVNLRWPFYIGITGAAGGALIALFDIK 420
 DB 361 FFLAKSEKLGAGASGVSAVLGITPAIFGVNLRWPFYIGITGAAGGALIALFDIK 420
 QY 421 AVAALGAGFLGVVYSIDAPDMVFLVCAVTVFVIAFGAAIYAGLYLVRRNGSIDPDATAAP 480
 DB 421 AVAALGAGFLGVVYSIDAPDMVFLVCAVTVFVIAFGAAIYAGLYLVRRNGSIDPDATAAP 480
 QY 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATIALSSVDAMPASGKLGSSVAIVPTKGL 540
 DB 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATIALSSVDAMPASGKLGSSVAIVPTKGL 540
 QY 541 VSPVSGKIVVAFPSGHAFVTRKAEKDSNVNIDILMHIGFDTVNLNGHFNPLKQSGDEKA 600
 DB 541 VSPVSGKIVVAFPSGHAFVTRKAEKDSNVNIDILMHIGFDTVNLNGHFNPLKQSGDEKA 600
 QY 601 GELLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPNVYGLGEIPAGANLLNVAKKEAVPAT 660
 DB 601 GELLCEFDIDAIAKAAGYEVTTPIVNSYKKTGPNVYGLGEIPAGANLLNVAKKEAVPAT 660
 QY 661 P 661
 DB 661 P 661

RESULT 2

AAAG92650
 ID AAG92650 standard; Protein; 661 AA.
 XX
 AC AAG92650;
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 6404.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH67869.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS Claim 17; SEQ ID NO: 6404; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 661 AA;
 Query Match 99.0%; Score 3310; DB 22; Length 661;
 Best Local Similarity 98.9%; Pred. No. 0;
 Matches 654; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MDHKLQAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60
 DB 1 MDHKLQAQRILRDIGGEDNIVAACHCATRLRLVLDKTDVDRQSLDDDDPDLKGTFTGGM 60
 QY 61 FOIIVGPGDVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120
 DB 61 FOIIVGPGDVHVFKELDATSKDIAVSTEQIKDVVANNANFSAVKVLADIFVPLIPI 120
 QY 121 LVGGGLLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVPGFTATK 180
 DB 121 LVGGGLLMAINNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLVPGFTATK 180
 QY 181 RFGCNFLGAGIGMAMVFTLVNGYDVAATMAGEMPMSLFGDVAQAGYOGTVLPVLY 240
 DB 181 RFGCNFLGAGIGMAMVFTLVNGYDVAATMAGEMPMSLFGDVAQAGYOGTVLPVLY 240

Db 181 RFGNEFLCAGIGMVFPSLVNGYDVAATMAAGEMPMSLFGLDVAAGYQGTVPVLV 240
Qy 241 VSNLATIEKFLKRLMGADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLOGLY 300
Db 241 VSNLATIEKFLKRLKGTADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLOGLY 300
Qy 301 DFGPGVGLLFGVYSPVITGLHQSFPPFIELEFNQGGSFIFATASMANIAQGAACLA 360
Db 301 DFGPGVGLLFGVYSPVITGLHQSFPPFIELEFNQGGSFIFATASMANIAQGAACLA 360
Qy 361 FFLAKSEKLGKLAGASVSNAVLTGITEPAIFGVNLRWRPFYIGIGTAAIGGALLIALFDIK 420
Db 361 FFLAKSEKLGKLAGASVSNAVLTGITEPAIFGVNLRWRPFYIGIGTAAIGGALLIALFNK 420
Qy 421 AVALGAAGFGLGVYSDAPDMVFLVCVVTFFVIAFGAAIAYGLYLVRNGSIDPDATAAP 480
Db 421 AVALGAAGFGLGVYSDAPDMVFLVCVVTFFVIAFGAAIAYGLYLVRNGSIDPDATAAP 480
Qy 481 VPAGTTKAEAEAPAEFSNDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGOL 540
Db 481 VPAGTTKAEAEAPAEFSNDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGOL 540
Qy 541 VSPVSGKIVVAFPSGHAFVTRKAEDEGSDNVDILMHIGFDTVNLNGTHFNPLKQGDVKA 600
Db 541 VSPVSGKIVVAFPSGHAFVTRKAEDEGSDNVDILMHIGFDTVNLNGTHFNPLKQGDVKA 600
Qy 601 GELLCEFDIDAIAKAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPAT 660
Db 601 GELLCEFDIDAIAKAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPAT 660
Qy 661 P 661
Db 661 P 661

RESULT 3

AAAB66707
ID AAB66707 standard; protein; 468 AA.

XX AC AAB66707;

XX DT 09-APR-2001 (first entry)

XX DE C-glutamic phosphoenolpyruvate protein #1.

XX KW Phosphoenolpyruvate; sugar phosphotransferase system; PTS.

XX OS Corynebacterium glutamicum.

XX PN W0200102583-A2.

XX PD 11-JAN-2001.

XX PF 27-JUN-2000; 2000WO-IB00973.

XX PR 01-JUL-1999; 99US-0142691.

XX PR 23-AUG-1999; 99US-0150310.

XX PR 03-SEP-1999; 99DE-1042095.

XX PR 03-SEP-1999; 99DE-1042097.

XX PA (BADI) BASF AG.

XX PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

XX XX WPI; 2001-080989/09.

XX XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate;
PT sugar phosphotransferase system proteins or their portions, useful for
PT typing or identifying C. glutamicum or related bacteria, and as markers
PT for transformation -

XX PS Claim 4; Page 101-102; 144pp; English.

XX PS

CC The present invention relates to Corynebacterium glutamicum
CC phosphoenolpyruvate; sugar phosphotransferase system (PTS) proteins.
CC The PTS nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria, the typing or identification of C. glutamicum or
CC related bacteria, as reference points for mapping C. glutamicum genome,
CC and as markers for transformation.

XX Sequence 468 AA;

Qy Query Match 69.8%; Score 2332; DB 22; Length 468;
Best Local Similarity 98.5%; Pred. No. 7.2e-223;
Matches 461; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 194 MAMVPTLVNGYDVAATMAAGEMPMSLFGLDVAAGYQGTVPVLVVSNIATIEKFLH 253
Db 1 MAMVFPVLNGYDVAATMAAGEMPMSLFGLDVAAGYQGTVPVLVVSNIATIEKFLH 60
Qy 254 KRLMGADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLOGLYDFGPGVGLLFG 313
Db 61 KRLMGADFLITPVLTLLTGTFTTAIGPAMRWGDLAHGLOGLYDFGPGVGLLFG 120
Qy 314 VSPVITGLHQSFPPFIELEFNQGGSFIFATASMANIAQGAACLAFFLAKSEKLG 373
Db 121 VSPVITGLHQSFPPFIELEFNQGGSFIFATASMANIAQGAACLAFFLAKSEKLG 180
Qy 374 GASGVSAVLGTEPAIFGVNLRWRPFYIGIGTAAIGGALLIALFDIKAVAGAGFLGV 433
Db 181 GASGVSAVLGTEPAIFGVNLRWRPFYIGIGTAAIGGALLIALFDIKAVAGAGFLGV 240
Qy 434 SIDAPDMVFLVCVVTFFVIAFGAAIAYGLYLVRNGSIDPDATAAPVPAETKAEAP 493
Db 241 SIDAPDMVFLVCVVTFFVIAFGAAIAYGLYLVRNGSIDPDATAAPVPAETKAEAP 300
Qy 494 AEFSDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKITVAVFP 553
Db 301 AEFSDSTIIQAPLTGEATIALSSVSDAMFASGKLGSGVAIVPTKGQLVSPVSGKITVAVFP 360
Qy 554 SCHAFVTRKAEDEGSDNVDILMHIGFDTVNLNGTHFNPLKQGDVKAAGELCEFDIDA 613
Db 361 SCHAFVTRKAEDEGSDNVDILMHIGFDTVNLNGTHFNPLKQGDVKAAGELCEFDIDA 420
Qy 614 AAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPATP 661
Db 421 AAGYEVTPPIVVSNNKKTGPVNTYGLGEIEAGANLLNVAKEAVPATP 468

RESULT 4

ABG80325
ID ABG80325 standard; Protein; 468 AA.

XX AC ABG80325;

XX DT 15-NOV-2002 (first entry)

XX DE C. glutamicum metabolic pathway (MP) protein #5.

XX KW Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;
KW cofactor; nucleotide; nucleoside; trehalose; fine chemical production;
KW organic acid; non-proteinogenic amino acid; purine base; carbohydrate;
KW pyrimidine base; lipid; unsaturated fatty acid; diol; polyketide;
KW aromatic compound; food industry; animal feed; cosmetic industry;
KW pharmaceutical industry; enzyme.

XX OS Corynebacterium glutamicum ATCC 13032.

XX PN W0200251231-A1.

XX XX 04-JUL-2002.

XX PS 22-DEC-2000; 2000WO-EP13143.

XX PS

```

PR 22-DEC-2000; 2000WO-EP13143.
XX (BADI ) BASF AG.
XX Pompejus M, Kroeger B, Zelder O, Schroeder H;
XX WPI; 2002-643289/69.
XX N-PSDB; ABS65346.
XX New metabolic pathway genes of Corynebacterium glutamicum for producing
XX fine chemicals, e.g. lipids, (unsaturated fatty acids, vitamins,
XX cofactors or enzymes used in food, feed, cosmetics or pharmaceutical
XX industries -
XX Claim 18; Page 111-113; 176pp; English.
XX The present invention relates to the isolation of Corynebacterium
XX glutamicum metabolic pathway (MP) proteins, and the polynucleotide
XX sequences encoding them. The MP proteins are enzymes involved in
XX the metabolism of molecules important for the normal functioning
XX of cells (e.g. amino acids, vitamins, cofactors, nucleotides and
XX nucleosides, or trehalose). The polynucleotide sequences encoding
XX the MP proteins are useful for producing fine chemicals, particularly
XX organic acids, non-proteinogenic amino acids, purine and pyrimidine
XX bases, nucleosides, nucleotides, lipids, (unsaturated fatty acids,
XX diols, carbohydrates, aromatic compounds, vitamins, cofactors,
XX polyketides and enzymes). The fine chemicals are useful in the food,
XX animal feed, cosmetic or pharmaceutical industries. ABG80321-ABG80343
XX represent the C. glutamicum MP proteins of the invention.
XX Sequence 468 AA;
XX
XX Query Match 59.8%; Score 2332; DB 23; Length 468;
XX Best Local Similarity 98.5%; Pred. No. 7.2e-233;
XX Matches 461; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
XX
QY 194 MAMVFTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYGTGVLPLVWSWILATIEKFLH 253
DB 1 MAMVFTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYGTGVLPLVWSWILATIEKFLH 60
QY 254 KRLMGADFLITPVLLLTGTFTTAIGPAMRWGDLAHLGQLGIDFGPGVGLLFLG 313
DB 61 KRLMGADFLITPVLLLTGTFTTAIGPAMRWGDLAHLGQLGIDFGPGVGLLFLG 120
QY 314 VYSPVITGLHQSPFPIELFNQGGSFIFATASMANIAQGAACIAVFLAKSEKLGIA 373
DB 121 VYSPVITGLHQSPFPIELFNQGGSFIFATASMANIAQGAACIAVFLAKSEKLGIA 180
QY 374 GASGSAVLGITEPAIFGVNLRWPFYIGITAAIGGALIALFDIKAVALGAAGFLGV 433
DB 181 GASGSAVLGITEPAIFGVNLRWPFYIGITAAIGGALIALFNKAVALGAAGFLGV 240
QY 434 SIDAPDMWFLYCAVVTFFIAFGAAIAYGLYLVRNGSIDDPATAAPVAGTTKAEAP 493
DB 241 SIDAPDMWFLYCAVVTFFIAFGAAIAYGLYLVRNGSIDDPATAAPVAGTTKAEAP 300
QY 494 AEFSNDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKGOLVSPVSGKIWAFF 553
DB 301 AEFSNDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKGOLVSPVSGKIWAFF 360
QY 554 SGHAFVAVRKAEDGSNDILMHIGFDVTNLTNTHNPLKKQDEVKAGELICEFDIDA 613
DB 361 SGHAFVAVRKAEDGSNDILMHIGFDVTNLTNTHNPLKKQDEVKAGELICEFDIDA 420
QY 614 AAGYEYVTTPIVSNYKKTGPVNTYGLGEIAGANLLNVAKKEAVPATP 661
DB 421 AAGYEYVTTPIVSNYKKTGPVNTYGLGEIAGANLLNVAKKEAVPATP 468
XX
XX RESULT 5
XX AAB66708
XX ID AAB66708 standard; protein; 362 AA.
XX

```

```

AC AAB66708;
XX 09-APR-2001 (first entry)
XX C.glutamicum phosphoenolpyruvate protein #2.
XX Phosphoenolpyruvate; sugar phosphotransferase system; PTS.
XX Corynebacterium glutamicum.
XX WO200102583-A2.
XX 11-JAN-2001.
XX 27-JUN-2000; 2000WO-IB00973.
XX 01-JUL-1999; 99US-0142691.
XX 23-AUG-1999; 99US-0150310.
XX 03-SEP-1999; 99DE-1042095.
XX 03-SEP-1999; 99DE-1042097.
XX (BADI ) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-080989/09.
XX Corynebacterium glutamicum nucleic acids encoding phosphoenolpyruvate:
XX sugar phosphotransferase system proteins or their portions, useful for
XX typing or identifying C. glutamicum or related bacteria, and as markers
XX for transformation -
XX Claim 6; Page 104-106; 144pp; English.
XX The present invention relates to Corynebacterium glutamicum
XX phosphoenolpyruvate: sugar phosphotransferase system (PTS) proteins.
XX The PTS nucleic acids and proteins which can be used to produce fine
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria, the typing or identification of C. glutamicum or
XX related bacteria, as reference points for mapping C. glutamicum genome,
XX and as markers for transformation.
XX Sequence 362 AA;
XX
XX Query Match 54.0%; Score 1806; DB 22; Length 362;
XX Best Local Similarity 99.2%; Pred. No. 1.1e-170;
XX Matches 359; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
QY 300 YDFGPGVGLLFLGLVYSPVITGLHQSPFPIELFNQGGSFIFATASMANIAQGAACIA 359
DB 1 YDFGPGVGLLFLGLVYSPVITGLHQSPFPIELFNQGGSFIFATASMANIAQGAACIA 60.
QY 360 VFFLAKSEKLGLAGASVSVALGITEPAIFGVNLRWPFYIGITAAIGGALIALFDI 419
DB 61 VFFLAKSEKLGLAGASVSVALGITEPAIFGVNLRWPFYIGITAAIGGALIALFNI 120
QY 420 KAVALGAAGFLGVYSIDAPDMWFLYCAVVTFFIAFGAAIAYGLYLVRNGSIDDPATA 479
DB 121 KAVALGAAGFLGVYSIDAPDMWFLYCAVVTFFIAFGAAIAYGLYLVRNGSIDDPATA 180
QY 480 PVPAGTTKAEAPAEFSNDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKG 539
DB 181 PVPAGTTKAEAPAEFSNDSTIIQAPLTGEALSSVDAMFASGKLGSGVAIVPTKG 240
QY 540 LVSPVSGKIYVAFPSGHAFVAVRKAEDGSNDILMHIGFDVTNLTNTHNPLKKQDEVK 599
DB 241 LVSPVSGKIYVAFPSGHAFVAVRKAEDGSNDILMHIGFDVTNLTNTHNPLKKQDEVK 300
QY 600 AGEELICEFDIDAIIKAAGYEYVTTPIVSNYKKTGPVNTYGLGEIAGANLLNVAKKEAVPA 659
DB 301 AGEELICEFDIDAIIKAAGYEYVTTPIVSNYKKTGPVNTYGLGEIAGANLLNVAKKEAVPA 360

```

QY 660 TP 661
 II
 Db 361 TP 362

RESULT 6
 ABB47495
 ID ABB47495 standard; Protein; 617 AA.
 XX AC ABB47495;
 XX DT 05-FEB-2002 (first entry)
 XX DE Listeria monocytogenes protein #199.
 XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 XX KW vitamin B12; bacterial infection; disease.
 XX OS Listeria monocytogenes.
 XX PN WO200177335-A2.
 XX PD 18-OCT-2001.
 XX PF 11-APR-2001; 2001WO-FR01118.
 XX PR 11-APR-2000; 2000FR-0004629.
 XX PA (INSP) INST PASTEUR.
 XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain I, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX WPI: 2002-010914/01.
 XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 XX PT and prevention of Listeria and related bacterial infections, and
 XX PT related polypeptides -
 XX PS Claim 6; SEQ ID No 200; 192pp; French.
 XX CC The present invention relates to the genome sequence of Listeria
 XX CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 XX CC it are useful for selecting probes and primers for detecting genes in L.
 XX CC monocytogenes and related organisms, and for studying genetic
 XX CC polymorphisms and other genomes. The present sequence is a protein
 XX CC encoded by the genome sequence of the present invention. Proteins
 XX CC expressed from the genome sequence are useful for raising specific
 XX CC antibodies, identification of L. monocytogenes and related organisms, and
 XX CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 XX CC B12. The genome sequence and proteins encoded by it are also useful for
 XX CC selecting compounds that regulate gene expression and cell replication
 XX CC and modulate L. monocytogenes-related diseases. In addition, the genome
 XX CC sequence and proteins encoded by it are useful in pharmaceutical and
 XX CC vaccines compositions for the treatment or prevention of infections by L.
 XX CC monocytogenes and related organisms.
 XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pat_sequences.

Sequence 617 AA;
 Query Match 28.6%; Score 957; DB 23; Length 617;
 Best Local Similarity 35.1%; Pred. No. 8.4e-86;
 Matches 225; Conservative 124; Mismatches 228; Indels 64; Gaps 17;

QY 1 MDHKLQRIILRDIGGEDNIVAAAHCAATRLRLVLKDKVDQSLDDDDPDLKGTFTGGM 60

Db 1 MDYQKLAKELIANVGGEENVRSVHCAATRLRFLKLVNKEKADKKQIESISGYSIVVENAGQ 60
 QY 61 FQITVG--PGDVHDV---FKELDDATSKDIAVSTEQKLDVVANNANFNSRAVKVLADIFV 115
 Db 61 LQVILGNVGDVYKALGSFTKLTDDGSEIAKGT---KD---SGNFSLKRAIDVISGIFT 114
 QY 116 PLIPLVGGGLMAINNVYLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAELPLVVG 175
 Db 115 PLGALAGGGMKGLLMLTT---FG-----WLTESGTYQILYAADSIFYELPLILA 165
 QY 176 FTATKRFPGNEFLGAGICGMAMVFTLVNGYDVAATMTAGEMPMSLFGDLVQAQGYGTV 235
 Db 166 YTAARKGANPPVATAACALVYPTMINLFNEGAHIFLQIP-----VLMYSFSV 217
 QY 236 LPVLVYVILATIEKFLHKLRLMGTRADFLITPVLTLTLTGLTFTFAIGPAMRWGDLIAHG 295
 Db 218 IPIILAVWFLSILERFLNSKIHEAAKTELTPMCLMLIVPLTLTFAFGPLGTFISQGLASG 277
 QY 296 LQGLYDFGPGVGLLFLGLVYSPITVTLGHQSPPIEL-ELFNQGSFFIFATASMANIAQG 354
 Db 278 YTFIYNLSPIVAGAFMGAFQVLVIFGIHMGFVPIMINNSRYGDRDTMIAMVGPSNFAQA 337
 QY 355 AACLAVFFLAKSEKLGKLAGASGVSAILGITEPAIFGVNLRRLRNPFIIGICTAAIGGALI 414
 Db 338 GASLGVFLKTKKPEVKAIAGSAALTGFFGITEPSIYGVTLKYKKPFIASITAGAIQAIV 397
 QY 415 ALFDIKAVALGAAGFLGVVSDADPMV---MFLVCVVTFVIAFGAAIAY-----GLYLV 466
 Db 398 -----GAAGSSGAANA-IPGILTLPIFGKGVGFIL--GIAYAILLSAIGTFF 444
 QY 467 RRGSIDPDATAAPVPAGTTKAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGK 526
 Db 445 ---GKDEKADGI---APTKEAKETGVE---AEVIVPIRGNVPLNEVKDEAFSAGL 494
 QY 527 LSGVAIVPTKQLVSPVSGKIVAPPSGHAPAVETKAEKDSNDVILMHIGEDTVNLNGT 586
 Db 495 LKGVAIVPQEGKLISPVNGTITETAPTCHAGIRNS---DKGVEILLHVGEDIVOLNKG 550
 QY 587 HFNPLKKQGDVEKAGELICEFDIDAIAKAAGYEVTTPVYVN 627
 Db 551 YFKLLVAQGDRLVVGQALLEFQLEAKADGYDITTPIVTN 591

RESULT 7
 AAU37874
 ID AAU37874 standard; Protein; 627 AA.
 XX AC AAU37874;
 XX DT 14-FEB-2002 (first entry)
 XX DE Streptococcus pneumoniae cellular proliferation protein #303.
 XX KW Antisense; prokaryotic cellular proliferation protein;
 XX KW antibiotic; antibacterial; drug design.
 XX OS Streptococcus pneumoniae.
 XX PN WO200170955-A2.
 XX PD 27-SEP-2001.
 XX PF 21-MAR-2001; 2001WO-US09180.
 XX PR 21-MAR-2000; 2000US-191078P.
 XX PR 23-MAY-2000; 2000US-206848P.
 XX PR 26-MAY-2000; 2000US-207727P.
 XX PR 23-OCT-2000; 2000US-242578P.
 XX PR 27-NOV-2000; 2000US-253625P.
 XX PR 22-DEC-2000; 2000US-257931P.
 XX PR 16-FEB-2001; 2001US-269308P.


```
QY 67 PGDVDHFKELDDATSKDIAYSTEQLKDVANNANWFSAVKVLADIFVPLIPILVGGGL 126
Db 68 -NDVPFIYNDP-TAVSGIEGVSKEAASAKSNQNVGVMTTAEITPIIPALIVGGL 125
QY 127 LMAINNVL--VAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTATKRFGG 184
Db 126 ILGFRNVLEGVHWSMLDKGTTITESSQFWAGVNHFWLPLGPAIFQLPVGITWSVRKMG 185
QY 185 NEFLGAGTGMVFPVTLVNGVDVAATMTAGEMP--MWSLFGLDVAQAGYQGTVPVLVVS 242
Db 186 SOILGIVLIGLVSPQLLNAYAVASTPAADIAANWVNFYTNRIQIQVAPALLAG 245
QY 243 WILATIERFLKRLMGTADELITPVLTLTGFTTFAIGPAMRWVGLAHGQGLYDF 302
Db 246 LLSVLEIFHWKHIPVISMIFVFLSLIPALIAHTVLGP-----IGWTIGQLSSVLA 301
QY 303 G--GPVG---GLLGLVYSPVITGLHQSFPPIELF--NOGGSFIEATASMANIAOGAA 356
Db 302 GLTGPVKWLFGLAIFGALYAPFVITGLHMTNAIDTQIADAGGTALWFMIALNSIAQGA 361
QY 357 CLAVEFLAK--SEKLKLAGASGVSAVLGITEPAIFGVNLRRLRWPFIYIGTAAIGGALIA 415
Db 362 VFAYFMHRHDEREAQVSLPATISAVLGVTPEALFGVNVKYYIPFVAGMTGSALAGLSV 421
QY 416 LFDIKAVALGAAGFLGVSIDAPDMVP-----LYCAVVTFTVAFGAATAYGLLYVRNGS 471
Db 422 TPNVTAASITIGGLPGIISIQOYMLPFAGTMLVAIVVPMLLTF-----FFRKAGL 472
QY 472 IDPDATAVPVAGTAKAEAE---APAEFSNDSTI-----TOAPLTGEATALSVDAM 521
Db 473 FTK-----TEGDTNLQAEFAQEAEPVNEHEPVELTSVEIISPLTGQVKLSQATDPI 525
QY 522 PASGLKGSVAIVPTKQGLVSPYSKIVVAFPSGHAFVTRKAEDSGNVDDILMHIGFTV 581
Db 526 PASGVMGQGLVTEPSQGLTSPVNTVTLFTPKHAIGT--VSDEG--VELLIHIGMDTV 581
QY 582 NLNGTHFNPLKKQGDVKAAGELLCEFDIDAIAKAGVEVTTPIWVSN---YKKTGPVNTYG 638
Db 582 GLDGKGFESLVQGDHVTYVQQLIRFDMVDVIRAGLVETPTVIINQDAYTATIP-GIYP 640
QY 639 LGEIEAGANLL 649
Db 641 T-TIQAGASLM 650
```

RESULT 9
ABU02378
ID ABU02378 standard; Protein; 655 AA.

```
AC ABU02378;
XX
XX
XX
XX 11-FEB-2003 (first entry)
XX
XX DE S. pneumoniae type 4 strain protein from coding region #1956.
XX
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;
XX KW ear infection; antiinflammatory; antibacterial; immunostimulant;
XX KW auditory; respiratory; gene therapy; vaccine.
XX
XX OS Streptococcus pneumoniae type 4 strain.
XX
XX PN W0200277021-A2.
XX
XX PD 03-OCT-2002.
XX
XX PF 27-MAR-2002; 2002WO-IB02163.
XX
XX PR 27-MAR-2001; 2001GB-0007658.
XX
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
```

PI Masignani V, Tettelin H, Fraser C;
DR WPI: 2003-040579/03.
DR N-PSDB; ABX07668.

XX
PT New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection

XX Claim 1; SEQ ID No 3912; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABG56454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 655 AA;

Query Match 28.4%; Score 949; DB 24; Length 655;
Best Local Similarity 34.9%; Pred No. 5.8e-85;
Matches 234; Conservative 128; Mismatches 233; Indels 56; Gaps 19;

```
QY 7 AQRILRDIGEDNIVAAAHCAIRLRLVLDKTDVDRSLDDDDPKLGTFTTGMFQIIVG 66
Db 8 AKDLLQAIGGKENVTAETHCATRMRFVLGDDSKANVAIESIPAVKGTFTNAGOFVIIG 67
QY 67 PGDVDHFKELDDATSKDIAYSTEQLKDVANNANWFSAVKVLADIFVPLIPILVGGGL 126
Db 68 -NDVPFIYNDP-TAVSGIEGVSKEAASAKSNQNVGVMTTAEITPIIPALIVGGL 125
QY 127 LMAINNVL--VAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTATKRFGG 184
Db 126 ILGFRNVLEGVHWSMLDKGTTITESSQFWAGVNHFWLPLGPAIFQLPVGITWSVRKMG 185
QY 185 NEFLGAGTGMVFPVTLVNGVDVAATMTAGEMP--MWSLFGLDVAQAGYQGTVPVLVVS 242
Db 186 SOILGIVLIGLVSPQLLNAYAVASTPAADIAANWVNFYTNRIQIQVAPALLAG 245
QY 243 WILATIERFLKRLMGTADELITPVLTLTGFTTFAIGPAMRWVGLAHGQGLYDF 302
Db 246 LLSVLEIFHWKHIPVISMIFVFLSLIPALIAHTVLGP-----IGWTIGQLSSVLA 301
QY 303 G--GPVG---GLLGLVYSPVITGLHQSFPPIELF--NOGGSFIEATASMANIAOGAA 356
Db 302 GLTGPVKWLFGLAIFGALYAPVITGLHMTNAIDTQIADAGGTALWFMIALNSIAQGA 361
```


QY 357 CLAVFLAK-SKLLKGLACASGVSAVLGITEPAIFGVNLRWPYIIGTAAIGALIA 415
 Db 362 VFAYFMHRRDREQAQVSLPATISAYLGVTPEALFGVNVKYYIPFVAGMTGSAAGMLSV 421
 QY 416 LFDIKAVALLGAAGFLGVVSIDAPDMVF-----LVCAVTFVIAFGAAIAYGLYLVRNGS 471
 Db 422 TENVTAASIGIGLPGILSIQOYMLPFPAGTMLVAIVPMLLTF-----PPRKAGL 472
 QY 472 IDPDATAAPVPACTTKAEAE-----APAEFSDSTI-----IQAPLTGEATLSSVSAM 521
 Db 473 FTK-----TEGDNLQAEFVAQEAEEFVNEPELVLSVELISPLTGQVKELSQATDPI 525
 QY 522 FASGLKSGVAIVPTKQGLVSPVSGKIVVAFPSGHAFVARTKAEDGSNVDMHGTGDTV 581
 Db 526 FASGVMGQGLVIERPSGELTSPVNGTIVTLFPTKHAIGI--VSDG--VELLIHIGMTV 581
 QY 582 NLNGHFNPLKKGDEVKAGELLCBFDIDAIAKAAGYEVTTPIVSN---YKKTGPVNTYG 638
 Db 582 GLDGKGFSLVVGQDHVTVGQQLIRFDMVDVIRKAGLVTTETPVITNQDAYTATIP-GTYP 640
 QY 639 LGEIEAGANLL 649
 Db 641 T-TIQAGASIM 650

RESULT 10

ABP27216
 ID ABP27216 standard; Protein; 620 AA.

XX AC ABP27216;
 XX DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 3508.

XX KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus pyogenes.

XX PN WO200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB04789.

XX PR 27-OCT-2000; 2000GB-0026333.

XX PR 24-NOV-2000; 2000GB-0028727.

XX PR 07-MAR-2001; 2001GB-0005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX PI Tettelin H;

XX DR WP1; 2002-352536/38.

XX DR N-PSDB; ABN67847.

XX PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -

XX PS Claim 1; Page 3514; 4525pp; English.

XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX SQ Sequence 620 AA;

Query Match 28.0%; Score 935.5; DB 23; Length 620;
 Best Local Similarity 34.7%; Pred. No. 1.2e-83;
 Matches 226; Conservative 108; Mismatches 273; Indels 45; Gaps 11;

QY 1 MDKDLAQRILRDIGGEDNIVAAAHCAIRLRLVLKDKVDQRSLDDDDPOLKFTFGGM 60
 Db 1 MNRQIAAEVIEALGGRENVRSVAHCAIRLRLVMDYDEKIDKEAAIDKVGAFNSGQ 60
 QY 61 FOIIVGPDVDHVFKELDATSKDIIV-----STEOLKDVVANNANWFSRAVKVLADIFV 115
 Db 61 YQMIFGTGTYNNIYDEV-----VALGLPTSSTSEKAEAGKGNIFQRAIRTFGDFV 113
 QY 116 PLIPILVGGLLMAINNVILVQA-----DLFGPSLVEMFPQISGVAEMINIMASAPFALP 171
 Db 114 PIIPAIIVATGLFMGVRG-LVTQPAIMDLFGVHEGENFLMYT-----RLTDTAFVYLP 166
 QY 172 VLVGFTATKRFNGEFLGAGIGAMVFTLVNGDVAATMTAGEMWMSLFLGDLVQAQY 231
 Db 167 ALVAMSAFRVGGNPPIGIVGLMLVSNELPNWVVA--SGGDVKLTTFEGF-VPVVG 222
 QY 232 QGTVLPLVIVSMILATIEKFLHKLMTADFLTPVLTLLLTGLTFLTIALGPMRWGDL 291
 Db 223 QGTVLPAFFVGLVGALEKLEKLVKVPALDLVTPFTFAIMSTLGLFVGPVPHSLNL 282
 QY 292 LAHGLQGLYDFGPGVGLLGLVYSPVITGLHQSPFPIELF-NOGGSFIFATASMAN 350
 Db 283 VLAGTQAVLHLPFGIAGLIVGGIQQLIVTGIHIFNLEAQLIANTGKDPFNAYLTAAT 342
 QY 351 IAQGAACLAFFLAKSEKLGAGASVAVLGITEPAIFGVNLRWPFYIGITGAIG 410
 Db 343 AAQAGATLAVAVTKTKSLKGLAPFTLSALLGLTEPAIFGVNLRPKVFSVGLIGGALG 402
 QY 411 GALLALFDIKAVALGAAGFLGVSIDAPDMVMFLVCAVTFVIAFGAAIAYGLYLVRNG 470
 Db 403 GWAGLFGIAGTGFGIIVLPGLTLLYLNQQLLYLTMVLGLVAFATAYTWGY----- 455
 QY 471 SIDPDATAAPVPAGTTKAEAPAEFNSDSTIIQAPLTGEAIALSSVSDAMFASGLKGS 530
 Db 456 ---QDRETLPLPAVEVDQADQALABE---TLVSPNLGTVDLSVSDPVFSSGANGOG 509
 QY 531 VAIPTKGQLVSPVSGKIIVAFPSGHAFVARTKAEDGSNVDMHIGFDTVNLNGTHNP 590
 Db 510 LAIPEDNTLSPVDGKRVIEIVFETGHAYTIS-----SQGAELVLIHIGIDESMAGFES 565
 QY 591 LKQGDDEVKAGELICEFDIDAIAKAAGYEVTTPVWSNKKYTGPNVTYGLGEI 642
 Db 566 LVAVGQAVKKGDLGLGHFDPKSIABAGLDDTTMTIVSNIAQYQSVYDILAAQGHV 617

RESULT 11

ABP25654
 ID ABP25654 standard; Protein; 676 AA.

XX AC ABP25654;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polypeptide SEQ ID NO 484.

CC The present invention relates to the genome sequence of *Listeria*
CC monocytes EGB-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in *L.*
CC monocytes and related organisms, and for studying genetic
CC polymorphisms and related genomes. The present invention is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies. Identification of *L.* monocytes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate *L.* monocytes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by *L.*
CC monocytes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 634 AA;
Query Match 27.7%; Score 924.5; DB 23; Length 634;
Best Local Similarity 32.7%; Pred. No. 1.5e-82;
Matches 217; Conservative 126; Mismatches 229; Indels 91; Gaps 16;
QY 1 MDHKLDAORILRDIGEDNIVAAAHCAATRLRLVLKDTQKDVDRQSLDDPDLKGTFTGGM 60
DB 1 MKYEQLAKDLKNVGGKENSINVFHCITLRFKLDENIANTEKTEKLDGVISVKSQGG 60
QY 61 FOIIVPGDGDHVF-----ELDDATSKDIAVSTEQLDKDVANNANWFSRAVKVL 110
DB 61 YQVVG-NEVPDVFRAVLVGGISAEDEGSS---APAT-----GNIFREFDNI 106
QY 111 ADIFVPLIPILVGGILLMAINVLVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAL 170
DB 107 SGVFTPVGLAATGMKGFAMFVA---FG-----WITVTSQTYQLLYAIGDCLFYFF 157
QY 171 PVLVGTAKRFGNEFLGAGTGMWVPTLVNGYDVAATMTAGMPWMSLF----- 222
DB 158 PFLGYTAKKFGGNIFFGMAIGGALVPTL-----AGTAGD-PLYTLFAGTIFRSP 209
QY 223 -----GLDVAQAGQGTVPVLVSVWILATIEKFLHKLMTADFLITPVLTLTLTGL 276
DB 210 IHVTFGLIPVILMSYASSIPIVATYFGSKVEKFKIIPDVIKTVFVVPFCTLLIVPI 269
QY 277 TTAIGPAMRWGDLAHGLQGLYDFGPGVGLLEGLVSPVINGLHQSPPIEL-ELF 335
DB 270 TTVIGPIATWAGQLLAGTIWYNLSPITAGLILGGFQVVFGLHGLVFPVAINLT 329
QY 336 NOGGSFIFATASMANIAQGAACLAFFFLAKSEKRLAGASGVSAVLGITEPAIFGVNLR 395
DB 330 VLGHDPILAMTFGAFPAQGAVALVFFKSRKKIKSLSPATISGIFGVTPEPAIYGVTL 389
QY 396 LRWPFVIGTAAIGGALLALFDIKAVAGAGFLGVSIDAP-----DMYMLVCAV 449
DB 390 LKPFIMSCIAAGGIGIIFGAGSYINGGIFGLPNFFRPGSGISGEFVWVIAI 449
QY 450 TFVIAFGAAIAYGL-----YLVRNGSIDPDATAAPVPAGTAKAPAEPAEFSNDSTIIQ 504
DB 450 SFILGILYVVGKPADVWVEQSNTVE-----GETLIBRET-----IP 489
QY 505 APTGGAIALSSVSDAMFAGSKLGSCVAIVPKGQLVSPVSKIVVAPPSGHAFVRPKA 564
DB 490 APVPGVITADYKDEAFSSGALGKVAIPTVGRVYVAPAGTIVTTFPTGHAIGITTK- 548
QY 565 EDGSNDVILMHIGHFDVNLNGHFNPLKKGQDEVKAGELLCEPFDIAKAAQYEVNTPIV 624
DB 549 -DGA--EVLHIGNDIVQLEGKFTAHVQKQGVIEKQGLLTEDIEGIRKAGDYVTPV 605
QY 625 VSN 627
DB 606 VTN 608

RESULT 13
ABP25655
ID ABP25655 standard; Protein; 674 AA.
XX AC ABP25655;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 486.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352535/38.
XX DR N-PSDB; ABN66286.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX Claim 1; Page 3203; 4525pp; English.
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN6044-ABN71536 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX Sequence 674 AA;
SQ

Query Match 27.5%; Score 920; DB 23; Length 674;
Best Local Similarity 33.6%; Pred. No. 4.7e-82;
Matches 222; Conservative 122; Mismatches 256; Indels 60; Gaps 15;
QY 7 AQRILRDIGEDNIVAAAHCAATRLRLVLKDTQKDVDRQSLDDPDLKGTFTGGMFQIIVG 66
DB 8 AKSLTALGGENIKVYVTECHTRFVYLDNNDKANKVEIKISVVKGTFFNAGQFQIIVG 67
QY 67 PGDGDHVFKELDATSKDIAVSTEQLDKDVANNANWFSRAVKVLADIVPLIPILVGGGL 126
DB 68 -NDVP-VFYNDFTAVSSIEGVSKEAASAKSNQALQKRVMTMLAEITPTIPIIVGGL 125

RESULT 15
ABB49923
ID ABB49923 standard; Protein; 617 AA.
XX
XX
AC ABB49923;
XX
XX 05-FEB-2002 (first entry)
DT
XX
DE Listeria monocytogenes protein #2627.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Deboux P;
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
XX WPI; 2002-010914/01.
XX
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides
XX
XX Claim 6; SEQ ID NO 2628; 192pp; French.
XX
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccine compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 617 AA;
SQ
Query Match 25.7%; Score 859; DB 23; Length 617;
Best Local Similarity 30.3%; Pred. No. 4.8e-76;
Matches 202; Conservative 140; Mismatches 261; Indels 64; Gaps 14;
QY 1 MDHDLAQRILDIGGEDNVAACATRLRLVLRKDKVDROSLDDDPDLKGTFTGCM 60
DB 1 MDYNQLAKEILQAVGGKNNVNEVHCITRLRFLKQKQVNEKQLKGLDKVMGTNVAQG 60

QY 61 FQIIVGPGVDHVPKELDDAT-----SKDIATVSTEQLKDVVANNANWFSRAVKLADIFVP 116
DB 61 FQVIIG-NDVKVFDALAAENPAWKNKTTNKTRQKI-----KGFSEIFALSGVFAP 115
QY 117 LIPILVGGGLLMAINNVVAQDLFGPQSLVEMEPQISGVA-----EMINLMASAPFAEL 170
DB 116 ILPAIAGAGLI-----KGFMAFVSLGWLATDTETRYILLAIGDGVFEL 160
QY 171 FVLVGTATRRFGNREFLAGIGAMVFPPLVNGYOVAAATMTAGETPMNSLFGLDVAQAG 230
DB 161 PILVAVSAARYFRANMFVALGIGALLP-----DLTALLSAGTTP--HFICLPVPT 212
QY 231 YQGTFLPVLVVSMLATIEKFLHKLMTADFLITPVLTLTLGLTFTFAIGPAMRWVD 290
DB 213 YAYSVIPILLAIWMSYVEKWDRIPTSLKLLFVPLITMFIIVPLTLVVGPLGTFVGD 272
QY 291 LLAHGLQGLYDFGPGVGLLFGLVYSPVITGLHQSPPIEL-ELFNQGGSFIFATASMA 349
DB 273 GVSIGINLLNNGGAIGGILIGGAMAIIVMTGHHYAIPEVVISNLAKYGYDKFLPITYIS 332
QY 350 NIAOGAAFLAVEFLAKSEKLGLAGASGSAVLGITETPAIFGVNLRNLPFYIGITAAI 409
DB 333 NMSQAGATFGVFFRAKDKKLSLAFSGLTALMGVTEPAMYGINVYKRFPMASLIGGAA 392
QY 410 GGALIALFDIKAVAL-GAAGFLGVVSDAPDMVNFVCAVVFVFIAPGAIAIYGLILVRR 458
DB 393 GGGFAMFGVKAYVLTGNGGIPGLVGDTFVYALIAMALAFIVALIFSIFI 448
QY 469 NSGIDPDATAPVPAGTTKAEAPAFSNDSTIIQAPLAGEATLSSVSDAMFASGKL 528
DB 449 EQMVEATFVAEKIRAGT-----ELIQAPVTIGELVKMSQVNDTTFADGIMG 493
QY 529 SGVAIVPTKGOLVSPVSGKIVVAPPSGHAFAVTRKAEDGNSVDILMHIGFDVNLNTHF 588
DB 494 KSIKPNKNEGKLYAPANGTIISLFTKHAJAM--KSDNGA--EILLHVGIDTVKLDGNYF 549
QY 589 NPLKKQGDVYKAGELLCEFDIDAIAAGYEVVTTPIVVSNNYKKGTPVNTYIGIETEAGANL 648
DB 550 TAHVATGDVWEGQGLLVTFDMEKI-ABKYDTTMMVTNTNEYAVVEAKENGIVTKGNQV 608
QY 649 LNVAKKE 655
DB 609 MELRSEQ 615

Search completed: October 1, 2003, 05:01:40
Job time : 63 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 04:31:10 ; Search time 68 Seconds
(without alignments)
2508.423 Million cell updates/sec

Title: US-10-019-284A-2

Perfect score: 3342
Sequence: 1 MDHKLDAQRILRDIGEDNI.....IEAGANLLNVAKEAVPATP 661

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3310	99.0	661	16 Q8NMD6	Q8nmd6 corynebacte
2	1508	45.1	650	2 Q9S6S6	Q9s6s6 lactococcus
3	1264	37.8	534	2 Q93ML1	Q93ml1 lactobacill
4	1186	35.5	481	16 Q8CN82	Q8cn82 staphylococ
5	1183	35.4	480	16 Q8NV35	Q8nv35 staphylococ
6	1178	35.2	480	16 Q9R9Q0	Q9rq0 staphylococ
7	1098.5	32.9	627	16 Q9L8G6	Q9l8g6 clostridium
8	1093	32.7	479	16 Q9KLT8	Q9kl8 vibrio chol
9	1013.5	30.3	470	16 Q9KAS1	Q9kas1 bacillus ha
10	979.5	29.3	480	2 Q8GLN4	Q8gln4 pseudomonas
11	960	28.7	627	16 Q97PB8	Q97pb8 streptococ
12	957	28.6	617	16 Q8Y904	Q8y904 listeria mo
13	955.5	28.6	480	2 Q9F8X3	Q9f8x3 pseudomonas
14	956	28.6	627	16 Q8DNS8	Q8dns8 streptococ
15	951.5	28.5	627	16 Q8NZJ7	Q8nzj7 streptococ
16	949	28.4	655	16 Q97NW9	Q97nw9 streptococ

ALIGNMENTS

RESULT 1

Q8NMD6 PRELIMINARY; PRT: 661 AA.
ID Q8NMD6
AC Q8NMD6;
DT 01-OCT-2002 (TREMREL. 22, Created)
DT 01-OCT-2002 (TREMREL. 22, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Phosphotransferase system IIC components,
DE glucose/maltose/N-acetylglucosamine-specific (EC 2.7.1.69).
GN CGL2642.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF005282; BAC00036.1; -.
DR InterPro: IPR001127; PTS_EI1A.
DR InterPro: IPR001996; PTS_EI1B.
DR InterPro: IPR003352; PTS_EI1C.
DR InterPro: IPR003358; PTS_EI1A_1; 1.
DR Pfam: PF00367; PTS_EI1B; 1.
DR Pfam: PF02378; PTS_EI1C; 1.
DR ProDom: PD002243; PTS_EI1A; 1.
DR TIGRFAMs: TIGR00830; PTBA; 1.
DR PROSITE: PS01035; PTS_EI1B_CVS; 1.
KW Transferase,Complete proteome.
SQ SEQUENCE 661 AA; 69148 MW; C0E5E1C7833B9F74 CRC64;

Query Match 99.0%; Score 3310; DB 16; Length 661;
Best Local Similarity 98.9%; Pred. No. 1.4e-188;
Matches 654; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MDHKLDAQRILRDIGEDNIYAAAHGATRLRLVLKDTKDVRQSLDDDDPLKCTFTTGM 60
|||||

Db 1 MDHKDLAQRILKRDIGGEDNIVAAAHCAATRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 60
 QY 61 FOIIVPGDGDVHVFKELDGATSKDIADVSTEQKDVYVANNANWFSRAVKVLADIFVPLPI 120
 Db 61 FOIIVPGDGDVHVFKELDGATSKDIADVSTEQKDVYVANNANWFSRAVKVLADIFVPLPI 120
 QY 121 LVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLVLYGFTATK 180
 Db 121 LVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLVLYGFTATK 180
 QY 181 RFGGNEFLGAGIGMAKVFPTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTVPVLV 240
 Db 181 RFGGNEFLGAGIGMAKVFPTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTVPVLV 240
 QY 241 VSWILATIEKFLHKLKMGTADELITPVLTLTLTGTLTAIGPAMRWGDDLLAHGLQGLY 300
 Db 241 VSWILATIEKFLHKLKMGTADELITPVLTLTLTGTLTAIGPAMRWGDDLLAHGLQGLY 300
 QY 301 DFGGPGVGLLFGLVYSPVITGLHQSFPPIELELNFQGSFIFATASMANIAQAACLA 360
 Db 301 DFGGPGVGLLFGLVYSPVITGLHQSFPPIELELNFQGSFIFATASMANIAQAACLA 360
 QY 361 FFLAKSEKLGKLAGASGSAVLGITPEPAIFGVNLRWRPFYIGITAGGALIALFDIK 420
 Db 361 FFLAKSEKLGKLAGASGSAVLGITPEPAIFGVNLRWRPFYIGITAGGALIALFDIK 420
 QY 421 AVALGAAGFLGVYSDAPDMWFLVCVVTFVIAFGAAIAYGLVLRNGSIDPDATAP 480
 Db 421 AVALGAAGFLGVYSDAPDMWFLVCVVTFVIAFGAAIAYGLVLRNGSIDPDATAP 480
 QY 481 VPAGTTKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASKLGSVAIVPTKGL 540
 Db 481 VPAGTTKAEAPAEFSNDSTIIQAPLTGEATLSSVSDAMFASKLGSVAIVPTKGL 540
 QY 541 VSPVSKIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLNTHNPLKQGDVKA 600
 Db 541 VSPVSKIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLNTHNPLKQGDVKA 600
 QY 601 GELLCEFDIDAIAKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIAGANLLNVAKKEAVPAT 660
 Db 601 GELLCEFDIDAIAKAAGYEVTTPIVVSNYKKTGPVNTYGLGEIAGANLLNVAKKEAVPAT 660
 QY 661 P 661
 Db 661 P 661

RESULT 2

Q9S6S6 ID Q9S6S6 PRELIMINARY; PRT; 650 AA.
 AC Q9S6S6; 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Enzyme II sucrose protein (EC 2.7.1.69).
 GN SACH.
 OS Lactococcus lactis.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1358;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=99173919; PubMed=10074089;
 RA Luesink E.J., Marug J.D., Kuipers O.P., De Vos W.M.;
 ET "Characterization of the divergent sacB and sacR operons, involved
 in sucrose utilization by Lactococcus lactis.";
 RL J. Bacteriol. 181:1924-1926(1999).
 DR EMBL; Z97015; CAB09690.1; -;
 DR HSSP; P45618; 2GPR.
 DR InterPro; IPR001127; PTS_EIIB.
 DR InterPro; IPR001996; PTS_EIIB.
 DR InterPro; IPR003352; PTS_EIIC.
 DR Pfam; PF003358; PTS_EIIB_1; 1.
 Pfam; PF003367; PTS_EIIB; 1.

DR Pfam; PF02378; PTS_EIIC; 1.
 DR ProDom; PD002243; PTS_EIIB; 1.
 DR ProDom; PD001476; PTS_EIIB; 1.
 DR TIGRFAMS; TIGR00826; EIB_glc; 1.
 DR TIGRFAMS; TIGR00830; PTBA; 1.
 DR PROSITE; PS00371; PTS_EIIB_1; 1.
 DR PROSITE; PS01035; PTS_EIIB_Cys; 1.
 KW Transferase.
 SQ SEQUENCE 650 AA; 69636 MW; 250E26F7664D4204 CRC64;
 Query Match 45.1%; Score 1508; DB 2; Length 650;
 Best Local Similarity 46.9%; Pred. No. 1.5e-81;
 Matches 313; Conservative 123; Mismatches 192; Indels 40; Gaps 10;
 QY 1 MDHKDLAQRILKRDIGGEDNIVAAAHCAATRLRLVLKDTKDVDRQSLDDDDPDLKGFETGGM 60
 Db 1 MNHQAVERILNAV-GRDNTQGAHRCATRLRLVLKDTGVIDQERALNDPDKGFEAAGQ 59
 QY 61 FOIIVPGDGDVHVFKELDGATSKDIADVSTEQKDVYVANNANWFSRAVKVLADIFVPLI 118
 Db 60 YQIIVPGDVTNVEEFIKLTIGISEA-STADLKEIAGSKQKQNPVMAVLKLLSDIFVPLI 118
 QY 119 PILVGGGLLMAINNVLAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLVLYGFTA 178
 Db 119 PALVAGGLLMAINNALTAHLFATKSLVEMFPKMGADIVNTMSAAPPFTFPLIGYSA 178
 QY 179 TKRFGGNEFLGAGIGMAKVFPTLVNGYDVAATMTAGEMPMSLFGDLVAQAGYQGTVPV 238
 Db 179 TKRFGGPNPLCAVVMVMPGLINGYNVAEISNHTMTYWDIFGFKVAQAGYQGVLPV 238
 QY 239 LVVSWILATIEKFLHKLKMGTADELITPVLTLTLTGTLTAIGPAMRWGDDLLAHGLQ 298
 Db 239 IGVAFILAKLBRFFHKLNDADFTFPLLSVIITGLTFTTIVGPAIRFVNSGLTDGLV 298
 QY 299 LYDFGGPGVGLLFGLVYSPVITGLHQSFPPIELELNFQ-----GGSFIFATASMANIA 352
 Db 299 LYLTLGALGLMVFSGFYSAIVVTGLHQSFPPIELELNFQ-----GGSFIFATASMANIA 358
 QY 353 QGAACLAIVFTLAKSEKLGKLAGASGSAVLGITPEPAIFGVNLRWRPFYIGITAGTGA 412
 Db 359 QAGATFALLFTVTKIKTKALAAPAGVSAIGLITPEPAIFGVNLRWRPFYIGITAGTGA 418
 QY 413 LIALFDIKAVAGLGAAGFLGVYSDAPDMWFLVCVVTFVIAFGAAIAYGLVLYRNGSI 472
 Db 419 FMGLFHLVAVLSGSLGIFSIKAGYNLQFWISIFSLFIATVVTSTYGRMEAK--SI 476
 QY 473 DPDATAAPVPAGTTKAEAPAEFSNDSTIIQ-----APLTGEATLSSVSDAMFAS 524
 Db 477 TKE-----KNQONATQYQPEKVIIDPVKSGELLAPINGFVPLSDVSDPVFSK 525
 QY 525 GKLGSVAIVPTKGLVSPVSGKIVVAPPSGHAFVARTKAEDGSNVDILMHIGFTVNLN 584
 Db 526 EIMGKGIKPKSGELFSPADGEIILAYETGHAYCIKTK----NGGEVLLHIGIDTSMN 581
 QY 585 GTHFNPLKQGDVKAAGLCEFDIDAIAKAAGYEVTTPIVVSN---YKTKGPVNTYGLGE 641
 Db 582 GNGFTQNVKVGOKVAGDLGSLGDFKEIKKSGLDSTVIIVITNSASYNEILLPSEN--VD 639
 QY 642 IEAGANLL 649
 Db 640 IKVGEKIL 647

RESULT 3

Q93ML1 ID Q93ML1 PRELIMINARY; PRT; 534 AA.
 AC Q93ML1; 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Sucrose-specific enzyme II of the PTS (Fragment).
 GN SCRA.
 OS Lactobacillus sakei.

GN SE1959.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dudez A.-M., Chailiou S., Hissler L., Stentz R., Champomier-Verges M.,
 RT "Physical and genetic map of the Lactobacillus sakei 23K chromosome.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF010445; AAK92528.1;
 DR InterPro; IPR001127; PTS_EIIC.
 DR Pfam; PF00358; PTS_EIIC.1;
 DR ProDom; PD002243; PTS_EIIC.1;
 DR TIGRfam; TIGR00830; PTBA.1;
 DR PROSITE; PS00371; PTS_EIIC.1;
 FT NON-TER
 SQ SEQUENCE 534 AA; 56230 MW; 01804F9DE70C0089 CRC64;
 Query Match 37.8%; Score 1264; DB 2; Length 534;
 Best Local Similarity 47.6%; Pred. No. 3.6e-67;
 Matches 265; Conservative 99; Mismatches 155; Indels 38; Gaps 10;
 QY 114 FVPLIPILVGGGLLMAANNVLAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVL 173
 DB 1 FVPLIPALTAGGLLMAANNVLTGGLEGASIVQMFPQKGFPAEIVNMWSAPFTFLPIL 60
 QY 174 VGFTATKRFGEFAGIGMAMVFTPLVNGDYDAITMTAGEMPMSLFLGLDVAQAGYQG 233
 DB 1 IAFSATKRFGNPYLGAAGMMLVNPVNGYGVASISATGHTMYHVFGLNTAQAGYQG 120
 QY 234 TVLVVYVSWTILATIEKFLHRLMGTADELITPVLTLGLTGLFTFIAIGPAMRWVGLLA 293
 DB 121 QVPIVGVAFTLANLEKFFHKLNDADVFTFPMLSIIIGLFTLVGPALRIVSNGVT 180
 QY 294 HGLQGLDFGPGVGLLGLVYSPVITGLHQSPPPIELELF----NOGGSFIFATASMA 349
 DB 181 DSLWVAYQTLGAVGMGIFGLYSAILVTLGLHQSPPAETTLADIATKGSFIFPVAAMA 240
 QY 350 NIAQGAACLAFFFLAKSEKLGAGAGSVSAVLGITEPAFGVNLRLRWPFFYIGIGTAI 409
 DB 241 NIAQGAATFAFFVFTKKNQKSLTSAGISAMLGITPEALFGLKLFPPFGLIASGI 300
 QY 410 GGALIALFDIKAVAGAGFLGVSVISDAPDMVFLVCAVTFVIAFGAAIAYGLYLVRN 469
 DB 301 SSFIIGLHLVSVNGPAGIIGFTAIAPKSPISPMGAILISFVIAFVGTVLYGKKAMK-- 358
 QY 470 GSIDPDATAAPVPAGTAKAER--BAPA-----EFSNDSTIIQAPLTGEATLSVSDAMP 522
 DB 359 -----TTEEEIINEAPATPEYVERLQDEK--ISAPVTGRIVDLASVPDPVF 402
 QY 523 ASGKLGSVAIVPTKGLVSVSGKIVVAFPSGHAFVTRKAEDGSNNVDILMHIGFDTVN 582
 DB 403 ASEANGKGIAMPTSDQVLAFTVGTITANTGHAYGI--KSDDGA--EVLHIGLDTVN 458
 QY 583 LNGTHFNPLKQGGDEVKAGELLCDFDIDAIAKAAYEVVTPVIVSN---YKKTGPVNTYGL 639
 DB 459 LNGIGFEXIVQGGHVSSEGLLGHFDIDIKQAGLITPLTMTIVTINTAGYAQVDPLLVDK 518
 QY 640 GEIEAGANLNV-AKKE 655
 DB 519 AAWQ-GEETIQHAKKD 534
 RESULT 4
 Q8CN82 PRELIMINARY; PRT; 481 AA.
 ID Q8CN82
 AC Q8CN82
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE PTS system sucrose-specific IIBC component.

GN SE1959.
 OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF016750; AAO05600.1;
 DR Complete Proteome.
 KW SEQUENCE 481 AA; 51609 MW; 4472BFEB304F409 CRC64;
 Query Match 35.5%; Score 1186; DB 16; Length 481;
 Best Local Similarity 49.9%; Pred. No. 1.4e-62;
 Matches 235; Conservative 95; Mismatches 123; Indels 18; Gaps 6;
 QY 1 MDHKDLAORILDRIGGEDNIVAAACATRLRLVLKDTKDVDRSLDDDDDLKGTETGGM 60
 DB 1 MSYKSAEELKAIAGEENLDMACHATRLRLVLNDESKVDEDTLSNMVDVKGTFSTGGQ 60
 QY 61 FOIIVGPGVDVHFVKELDDATSKDIADVSTEQKLVVANNANWFSRAVKVLADIFVPLPI 120
 DB 61 YQIIIGSGTVNKVFNELEXITGKE-ASTTSEVKDKSKHNPQKFKVXKLSDFVPIIPA 119
 QY 121 LVGGGLLMAINNVLVAODLF-GPOSIVEMFPOISGVAEMINLMASAPFAFLPVLGFTAT 179
 DB 120 IVAGGLLMLGNLITFAKDLFYDGKSIIDVHSOFSLADMINIFANAPFTLLILIGFSA 179
 QY 180 KRFSGNEFLGAGIGMAMVFTPLVNGDYDAITMTAGE-MPMWSLFLGLDVAQAGYQGVLPV 238
 DB 180 KRFSGNPYLGAAAGMLVHPGLMSADYDFPKALEEKGAIHPWDVFGHLHNEVYQGVLEP 239
 QY 239 LVVSWTILATIEKFLHRLMGTADELITPVLTLGLTGLFTFIAIGPAMRWVGLLAHLOG 298
 DB 240 LVATYILATIEKFLHRLMGTADELITPVLTLGLTGLFTFIAIGPAMRWVGLLAHLOG 299
 QY 299 LYDFGPGVGLLGLVYSPVITGLHQSPPPIELELF----NOGGSFIFATASMANIAOG 354
 DB 300 LYDFGPGVGLLGLVYSPVITGLHQSPPPIELELF----NOGGSFIFATASMANIAOG 359
 QY 355 AACLAFFFLAK-SEKLGAGAGSVSAVLGITEPAFGVNLRLRWPFFYIGIGTAIIGGAL 413
 DB 360 GAALAAFFFIKQNKLGKLVASAGISALLGITEPAMFGVNLRLRWPFFYIGIGTAIIGGAL 419
 QY 414 IALFDIKAVAGAGFLGVSVISDAPD-----MMVFLVCAVTFVIA 454
 DB 420 ISFEKVAIALGTAGLPGFISINPHTAGMLHYLIGMLIAFVSVVTVLVLS 470
 RESULT 5
 Q8NV35 PRELIMINARY; PRT; 480 AA.
 ID Q8NV35
 AC Q8NV35
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE PTS system sucrose-specific IIBC component.
 GN SCRA OR MW2299.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
 RA Yamamoto K., Hiramatsu K.
 RT "Genome and virulence determinants of high virulence community-
 RT acquired MRSA.";
 RL Lancet 359:1819-1827(2002).
 DR EMBL; AP004830; BAB96164.1;


```
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR TIGRFAMs: TIGR00826; EIIB_glc; 1.
DR PROSITE: PS01035; PTS_EIIB-CYS; 1.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51218 MW; B3EBE81C08CODEDA CRC64;

Query Match 35.4%; Score 1183; DB 16; Length 480;
Best Local Similarity 50.4%; Pred. No. 2e-62;
Matches 242; Conservative 94; Mismatches 130; Indels 14; Gaps 7;

QY 1 MDHKLDAQIRLDIGEDNIVAAHCAATRLRLVLDKTDVDRQSLDDDDPDLKGFETGGM 60
DB 1 MNYKQSAEILNAGSEENLDAMAHCAATRLRLVLDNDESLVNEEALNNMNVVKGFTSGGQ 60
QY 61 FOIIVGPGDGVHVKELDDATSKDIAVSTEQLKDVVANNANWFSRAVKVLADIFVPLIPI 120
DB 61 YQIIGSGVNVKVFSELEKLTGKE-ASTTSEVKAQSAKNNPLQRFVKMLSDIFVPIIPA 119
QY 121 LVGGGLLMAINNVVAODL-FGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFTAT 179
DB 120 IVAGGLMLGNLLTAKDLFFSGKSLIDVYSQFAGLAEMINVFANAPTLLPILIGFSAA 179
QY 180 KRFGGNEFLGAGIGAMVPTLVNGYDVAATMTAGE-MPMWSLFLGLDVAQAQYQGTVLVP 238
DB 180 KRFGGNFFLGAALGMILVHPSLMSAYDFPKAVEAGKAIPYDWDVGLHINQVYGOQVLPM 239
QY 239 LVVSWILATIEKFLHKLMTGADFLITPVLTLLLTGFTLFTAIQPMRWVGDLLAHGLQG 298
DB 240 LVAAVILASTEKGLRKVIPTVLDNLLTLLSIFITAFITFVFGVPIITRQLGYWLSGDLTW 299
QY 299 LYDFGPGVGLLFLGLVYSPITVITGLHOSFPPIELELF-NOGGSFIFATASMANIAQG 354
DB 300 LYDFGGAIGLIPGLIYAFIVITGMHHSFIIVETTLIADATKIGSFIFPIATMSNAQG 359
QY 355 AACLAVFFLAK-SEKLGAGAGSVSAVLGITEPAIFGVNLRNLPFYIGTGAIGGAL 413
DB 360 GAAIAFAFFIKQNKKLGVASAGISALLGITTEPAMFGVNLKLRVPFPGAIVSGSIGSAY 419
QY 414 IALFDIKAVAGLGAAGFLGVVSIDA--PDMVMFLCAVVFVFAFGAIAIYGLVYVRNGS 471
DB 420 IAFKVKAIATAGTAGLPGFISINPVHAGWLHYFVGWMTISFII-----AITVTLILSKRKN 475

RESULT 6
Q99RQ0 PRELIMINARY; PRT; 480 AA.
AC Q99RQ0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE PTS system, sucrose-specific IIBC component.
GN SCRA OR SAV2377 OR SA1167.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
SEQUENCE FROM N.A.
SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RP MEDLINE=21311952; PubMed=11418146;
RA Kureda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";

Lancet 357:1225-1240(2001).
DR EMBL: AP003365; BAB58339.1; -.
DR EMBL: AP003137; BAB43469.1; -.
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR TIGRFAMs: TIGR00826; EIIB_glc; 1.
DR PROSITE: PS01035; PTS_EIIB-CYS; 1.
KW Complete proteome.
SQ SEQUENCE 480 AA; 51232 MW; D88607F6E0AF2E84 CRC64;

Query Match 35.2%; Score 1178; DB 16; Length 480;
Best Local Similarity 50.4%; Pred. No. 4e-62;
Matches 242; Conservative 94; Mismatches 130; Indels 14; Gaps 7;

QY 1 MDHKLDAQIRLDIGEDNIVAAHCAATRLRLVLDKTDVDRQSLDDDDPDLKGFETGGM 60
DB 1 MNYKQSAEILNAGSEENLDAMAHCAATRLRLVLDNDESLVNEEALNNMNVVKGFTSGGQ 60
QY 61 FOIIVGPGDGVHVKELDDATSKDIAVSTEQLKDVVANNANWFSRAVKVLADIFVPLIPI 120
DB 61 YQIIGSGVNVKVFSELEKLTGKE-ASTTSEVKAQSAKNNPLQRFVKMLSDIFVPIIPA 119
QY 121 LVGGGLLMAINNVVAODL-FGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGFTAT 179
DB 120 IVAGGLMLGNLLTAKDLFFSGKSLIDVYSQFAGLAEMINVFANAPTLLPILIGFSAA 179
QY 180 KRFGGNEFLGAGIGAMVPTLVNGYDVAATMTAGE-MPMWSLFLGLDVAQAQYQGTVLVP 238
DB 180 KRFGGNFFLGAALGMILVHPSLMSAYDFPKAVEAGKAIPYDWDVGLHINQVYGOQVLPM 239
QY 239 LVVSWILATIEKFLHKLMTGADFLITPVLTLLLTGFTLFTAIQPMRWVGDLLAHGLQG 298
DB 240 LVAAVILASTEKGLRKVIPTVLDNLLTLLSIFITAFITFVFGVPIITRQLGYWLSGDLTW 299
QY 299 LYDFGPGVGLLFLGLVYSPITVITGLHOSFPPIELELF-NOGGSFIFATASMANIAQG 354
DB 300 LYDFGGAIGLIPGLIYAFIVITGMHHSFIIVETTLIADATKIGSFIFPIATMSNAQG 359
QY 355 AACLAVFFLAK-SEKLGAGAGSVSAVLGITEPAIFGVNLRNLPFYIGTGAIGGAL 413
DB 360 GAAIAFAFFIKQNKKLGVASAGISALLGITTEPAMFGVNLKLRVPFPGAIVSGSIGSAY 419
QY 414 IALFDIKAVAGLGAAGFLGVVSIDA--PDMVMFLCAVVFVFAFGAIAIYGLVYVRNGS 471
DB 420 IAFKVKAIATAGTAGLPGFISINPVHAGWLHYFVGWMTISFII-----AITVTLILSKRKN 475

RESULT 7
Q9L8G6 PRELIMINARY; PRT; 627 AA.
AC Q9L8G6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ScrA (Fusion: PTS system, beta-glucosidase specific IIBAC
DE component).
GN SCRA OR CAC0423.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
SEQUENCE FROM N.A.
STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=20391269; PubMed=10937490;
RA Tangney M., Mitchell W.J.;
RA "Analysis of a catabolic operon for sucrose transport and metabolism
in Clostridium acetobutylicum ATCC 824.";
RL J. Mol. Microbiol. Biotechnol. 2:71-80(2000).
RN [2]
```



```
QY 356 AACLAFFFLAKSEKLAGASGVSAYLGTETPAIFGVNLRWLPFYIGTAAIGGALIA 415
Db 360 AALAVGVMTKEKLAGVAIPSGVTALLGTEPMEGVNLRYPPIAICGAALASAFIT 419
QY 416 LFDIKAVAGAGFGVGVSDAPDPMFVLCVAVTFVAFGMAI 459
Db 420 LFNVAQALGAAGLPGIITSINPQIGYIMGMAISFVAFAFTV 463

RESULT 9
Q9KAS1
ID Q9KAS1 PRELIMINARY; PRT; 470 AA.
AC Q9KAS1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE PTS system, trehalose-specific enzyme II, BC component.
GN BH2216.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001514; BAB05935.1; -
DR HSSP: P05053; 1IBA.
DR InterPro: IPR004719; PTS_IIC_glc.
DR InterPro: IPR001986; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR TIGRfams: TIGR00826; EIB_glc; 1.
DR TIGRfams: TIGR00852; PTS_glc; 1.
DR PROSITE: PS01035; PTS_EIIB_CYS; 1.
KW Complete proteome.
SQ SEQUENCE 470 AA; 50185 MW; 3CE67B1E9650F5B9 CRC64;
```

```
Query Match 30.3%; Score 1013.5; DB 16; Length 470;
Best Local Similarity 44.4%; Pred. No. 2.3e-52;
Matches 204; Conservative 90; Mismatches 150; Indels 15; Gaps 4;

QY 3 HKDLAQRILRDIGDENIVAAHCAHTRLRVLKTDVDRQSLDDDDPDLLKGFETGGMFQ 62
Db 4 YKKEVNAIVEAIGAGADNIQTATHTVRLRFVLRDEKGVDOEKLESLDIVKGSFSTNGQFQ 63
QY 63 IIIVGPGDVHVFREL-----DDATSKDIATVSTEQLKDVVANNANWFSRAVKVLADIFVP 116
Db 64 VIITGQGVTKYKELVAETGIGATKEDV-----KDAAKNTVQRAVKTADIFIP 116
QY 117 LPIVLVGGLLMAINNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVING 175
Db 117 ILPAIVTAGLGMGINILTAEIGFYDGAISYDIPQWADFAGIINLIANTAFVFLPLGLIG 176
QY 176 FTATKRRGGNEFLGAGIGMAWVPTLVNGVDVAATWTAGEMPWMSLFLGDVAQAGYQGV 235
Db 177 WSAVKRFGSGELGIVGLMLVHPDLLNANGYCAQLEGEIPTWNLFGLTIEQVGQGV 236
QY 236 LPVLVSWILATIEKFLHKLMTADFLTPVLTLLLTGFLTFIAGPAMRWVGDLLAHG 295
Db 237 LPILFSAWILAKIEIFURKRPVPSIQLLVAVPALLITGFIATAACIPITFTTGNGITNV 296
QY 296 LOGLYDFGGVGLLGLVSPVITGLHQSFPPIELEFNQ-GGSFIFATASMANIAQG 354
Db 297 FTSIFAAPVLVGGFLYLIYAPLVLTGMHHTFLAVDLQLLIGTIGGTFLLPILVLSIAQG 356
```

```
QY 355 AACLAFFFLAKSEKLAGASGVSAYLGTETPAIFGVNLRWLPFYIGTAAIGGALIA 414
Db 357 SAALAMFATRDKEKLGSLSSAVSAYLGTETPAMFGVNIRYKFPFVFCALISAIGGAFI 416
QY 415 ALFDIKAVAGAGFGVGVSDAPDPMFVLCVAVTFV1 453
Db 417 TVNGVLANSIGVGLPGIITSIQAGFGVGVFFIGWVIAFIL 455

RESULT 10
Q8GLN4
ID Q8GLN4 PRELIMINARY; PRT; 480 AA.
AC Q8GLN4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE TREP.
GN TREP.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O8r1-96;
RA Mavrodi D.V., de la Fuente L., Mavrodi O.V., Wellier D.M.,
RA Thomasow L.S.;
RT "Trehalose utilization operon of biocontrol strain Pseudomonas
RT fluorescens O8r1-96."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY131006; AAN06918.1; -
SQ SEQUENCE 480 AA; 51311 MW; B1F632B63042A10F CRC64;
```

```
Query Match 29.3%; Score 979.5; DB 2; Length 480;
Best Local Similarity 40.8%; Pred. No. 2.5e-50;
Matches 193; Conservative 110; Mismatches 161; Indels 9; Gaps 7;

QY 2 HKDLAQRILRDIGDENIVAAHCAHTRLRVLKTDVDRQSLDDDDPDLLKGFETGGMF 61
Db 4 DYSKIASLQSLGVDNIEQAHCYTRLRSLKDAARVDSATLQJDLVKGSFETGGLF 63
QY 62 QIIIVGPGDVHVFRELDDATSKDIATVSTEQLKDVVANNANWFSRAVKVLADIFVPLIPIL 121
Db 64 QVIVGPGEVKRYAALRELTGL-AAATVADVKRQAGKNGMQRVLRVSDVFPILPAL 122
QY 122 VGGGLMAINNVVAQDLF-GPQSLVEMFPQISGVAEMINLMASAPFAFLPVINGFTATK 180
Db 123 VIAGLLMGVNNLLGAKGFIAGQTLDDAYPHLDGWSLINLMANTSFVFLFALYGSAAK 182
QY 181 RFGGNEFLGAGIGMAWVPTLVNGVD---VAATWTAGEMPWMSLFG-LDVAQAGYQGVTL 236
Db 183 RFGSEILGIVGLMLVHPDLLNANGYKAVAGLEGQSLPYDFIEGWFRIEKVYQGOIL 242
QY 237 PVLVYVSWILATIEKFLHKLMTADFLTPVLTLLLTGFLTFIAGPAMRWVGDLLAHG 296
Db 243 PILLAAVYVMSVIERLRARVNAIQLLVIPITIVITGVLAALAIIGPVTRHLGLITEGM 302
QY 297 QGLYDFGPGVGLLGLVSPVITGLHQSFPPIELEFNQ-GGSFIFATASMANIAQGA 355
Db 303 VALFDLAPVIGVAGLPGLLYAPLVLTGMHMTFLAVDLQLISTQGGTFTFPMVIMNLAQGS 362
QY 356 ACLAVFFFLAKSEKLAGASGVSAYLGTETPAIFGVNLRWLPFYIGTAAIGGALIA 415
Db 363 AALAVFYVTRNARDKSMASSTSAISAYFGITEPAMFGVNIRYKFPFVFCALISAIGCIPLS 422
QY 416 LFDIKAVAGAGFGVGVSDAPDPMV-MFLYCAVTVFVIAFGAALAYGLVLR 467
Db 423 LNKIKASIAIGVGLPGFISI-VPDIPMFVGVMIAMSVFPALTCALSKMKIVR 474

RESULT 11
Q97PB8
```

ID	Q97PB8	PRELIMINARY;	PRT;	627 AA.
AC	Q97PB8;			
DT	01-OCT-2001 (TREMELrel. 18, Created)			
DT	01-OCT-2001 (TREMELrel. 18, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	PTS system IIABC components.			
GN	SP1722.			
OS	Streptococcus pneumoniae.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1313;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FIGR4;			
RC	MEDLINE=2137209; PubMed=11463916;			
RA	Tetelin S., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,			
RA	Peterlin S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,			
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,			
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,			
RA	Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,			
RA	McDonald L.A., Feldblyum T.V., Angluoli S., Dickinsohn T., Hickey E.K.,			
RA	Holt I.E., Jofstus B.J., Yang F., Smith H.O., Venter J.C., C.M.,			
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,			
RT	"Complete genome sequence of a virulent isolate of Streptococcus			
RT	pneumoniae."			
RL	Science 293:498-506(2001).			
RL	EMBL: AE007465; AAK75799.1; -			
DR	TIGR; SPI722;			
DR	InterPro; IPR001127; PTS_EIIA.			
DR	InterPro; IPR001996; PTS_EIIB.			
DR	InterPro; IPR003352; PTS_EIIC.			
DR	Pfam; PF00358; PTS_EIIA.1; 1.			
DR	Pfam; PF00367; PTS_EIIB; 1.			
DR	Pfam; PF02378; PTS_EIIC; 1.			
DR	ProDom; PD002243; PTS_EIIA; 1.			
DR	ProDom; PD001476; PTS_EIIB; 1.			
DR	TIGRFAMS; TIGR00826; EIIB-g1c; 1.			
DR	TIGRFAMS; TIGR00830; PTBA; 1.			
DR	PROSITE; PS00371; PTS_EIIA.1; 1.			
DR	PROSITE; PS01035; PTS_EIIB_CYS; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 627 AA; 65683 MW; 92980DD7B2E6B85F CRC64;			
Query Match	28.7%; Score 960;	DB 16;	Length 627;	
Best Local Similarity	33.3%;	Pred. No. 5e-49;		
Matches 222;	Conservative 141;	Mismatches 246;	Indels 58;	Gaps 14;
QY	1 MDKDLAQRILRDIGGEDNIVAAAHACATRLRLVLTQKDVDRQSLDDPDLKGTFTGGM 60			
DB	1 MNNQELAKKVIDALGGRENNVSAHACATRLRLVLTQKDVDRQSLDDPDLKGTFTGGM 60			
QY	61 FOIIVGPDVHVEKE---LDDATSKDIATVSTEQKLDVWANNKFSRAVKVLDLIFVPL 117			
DB	61 VOIIFGTGVNKMVDEVVVLGLPTS-----SKDDMKAEVAKQGNVFORAITFGDVFPI 115			
QY	118 IPILVGGGLLAINVLVAQILFGPSILVEMFPQTSVGAEMINLMASAPFAFLVPLVGT 177			
DB	116 IPVIVATGLFVGRGLFNALEM-----PLPGDFATYQIILDTAFILPGLVWVS 165			
QY	178 ATKRGGEFFLAGTGMAMWPTLVNGVDVAATMTAGEMPWLSFGLDVAAGVGTLP 237			
DB	166 TPRVFGGNPAVGIVLGMVLVSGSLFNWAVA---QGGEVTAMNFFGF-IPVVVGLQGSVLP 221			
QY	238 VLWVSKILATIEKFLKRLMGTAFLITPVLTLTGLTFTFAIGPAMRWVGDLLAHGLQ 297			
DB	222 ARIIGVGNKFKPAKVKVPPVDVILVFPVTVLLVNSILGLFVIGPVHVVNYILIAIK 281			
QY	298 GLYDFGGPVGGLFLGVSPVITVGLHSGFPPIELFNQGGSFTE-ATASMANTAQAA 356			
DB	282 AILSMFPLGGLGFLGVHQLVSGVHIFNLVQLLAADHANFNFAITTAAMTAQAA 341			
QY	357 CLAVFELAKSEKLGLAGASVAVLGIETPAIFGVNLRRLRPPFVIGTCTAIGGALLAL 416			

Db	342 TVAVGVTKNPKLKTALFPAALSAFLGITETPAIFGVNLRFRKPFLLSLIAGAIGGLASI 401			
QY	417 FDIKRAVALGAAGFLGVVSDAPDMV-----MFLVCVVTFVIAFGAALAYGLYLVR 468			
DB	402 -----LGLAGTNGIITIIIGTMLYVNGOLPOYLLMVAVSFALGALTVMFGY----- 449			
QY	469 NGSIDPDATAAPVPAGTTKAAE-APAEFSDNSTIIQAPLTGEGAIALLSSVDAMFASGKL 527			
DB	450 --EDEVDATAAKRAEVAEKEEVAAPALQNETLV--TPIVGDVVALADVNDVPFSSGAM 505			
QY	528 GSGVAIVPTKGQLYSPVSGKIIVVAPPSGHAFVARTKAEDGSNDVILMHIHIGFDIVNLNGTH 587			
DB	506 GOGIVVKSQGVVYAPADAESVIAPTGHAFGLKTR-----NGAEVLHVGIDITVSMNGDG 561			
QY	588 FNPLKKGQDEVKAGELLCFFDIDAIIKAAAGYEVTPPIVSVN---YKGTGPNVNTYGLGELEA 644			
DB	562 PETKVAQGNKVKAGDVLGTGFSNKIAAAGLDOTTMTIVITNRTGDSVAPVAT---GSAVK 618			
QY	645 GANLLNV 651			
DB	619 GDAVIEV 625			
RESULT 12				
Q8Y904	PRELIMINARY;	PRT;	617 AA.	
AC	Q8Y904;			
DT	01-MAR-2002 (TREMELrel. 20, Created)			
DT	01-MAR-2002 (TREMELrel. 20, Last sequence update)			
DT	01-MAR-2003 (TREMELrel. 23, Last annotation update)			
DE	Hypothetical protein lmo0738.			
GN	LMO0738.			
OS	Listeria monocytogenes.			
OC	Bacteria; Firmicutes; Bacilliales; Listeriaceae; Listeria.			
OX	NCBI_TaxID=1639;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BGD-e / Serovar 1/2a;			
RC	MEDLINE=21537279; PubMed=11679669;			
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,			
RA	Baquerio F., Berche P., Bloeker H., Brandt P., Chakraborty T.,			
RA	Charbit A., Chetouani F., Couve E., de Daruvar A., Durant L., Dusseret O.,			
RA	Doman E., Dominguez-Bernal G., Duchaud E., Durant L., Garrido P.,			
RA	Entian K.-D., Psihi H., Garcia-del Portillo F., Hain T., Hauf J., Jackson D.,			
RA	Gautier L., Goebel W., Gomez-Lopez N., Kuhn M., Kunst F., Kurapkat G.,			
RA	Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Nedjari H.,			
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari R.,			
RA	Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,			
RA	Rommel B., Rose M., Schlueter I., Simoes N., Tierrez A.,			
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,			
RT	"Comparative genomics of Listeria species."			
RL	Science 294:849-852(2001).			
RL	EMBL: AL591976; CAC98816.1; -			
DR	Listlist; LMO00738;			
DR	InterPro; IPR001127; PTS_EIIA.			
DR	InterPro; IPR001996; PTS_EIIB.			
DR	InterPro; IPR003352; PTS_EIIC.			
DR	Pfam; PF00358; PTS_EIIA.1; 1.			
DR	Pfam; PF00367; PTS_EIIB; 1.			
DR	Pfam; PF02378; PTS_EIIC; 1.			
DR	ProDom; PD002243; PTS_EIIA; 1.			
DR	ProDom; PD001476; PTS_EIIB; 1.			
DR	TIGRFAMS; TIGR00826; EIIB-g1c; 1.			
DR	TIGRFAMS; TIGR00830; PTBA; 1.			
DR	PROSITE; PS00371; PTS_EIIA.1; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 617 AA; 65447 MW; E7144ACDA7846A03 CRC64;			
Query Match	28.6%; Score 957;	DB 16;	Length 617;	
Best Local Similarity	35.1%;	Pred. No. 7.4e-49;		
Matches 225;	Conservative 124;	Mismatches 228;	Indels 64;	Gaps 17;
QY	1 MDKDLAQRILRDIGGEDNIVAAAHACATRLRLVLTQKDVDRQSLDDPDLKGTFTGGM 60			

DB 1 MDYQKLAKEILANVGGEENRVSWECAFLRFLKYNKEADKKQIESISGVISVVENAGQ 60
QY 61 FOIIIVG--PGVDVHV---FKLEDDATSKDIAVSTEQKDVVANNANWFSRAVKVLADIFV 115
DB 61 LQVIGNTGVGVYKALGSFTKLTDDGSEIAKT---KD---SDGNFSLKALDVISGIFT 114
QY 116 PLIPILVGGGLMANNVVAQDLFGPQSLVEMFPOISGVAEMINLMASAPFAFLPVLV 175
DB 115 PILGALAGGGLKGLIMLT---FG-----WLTSSGTQOILYAAADSDVFFLPLILIA 165
QY 176 FTATKRFEGNEFLGAGIGMAYFPTLVNGYDVAAMTAMTAGEMPMWSLFGDLVAQAQYQTV 235
DB 166 YTAARKEGANPPVAIAAGALVYPTMINLNFSGAHITFLQIP-----VVLMSYSFSV 217
QY 236 LPVLVSVTLATIEKFLHRLMGTDPLITPVLTLLTGLFTLFAIGPAMRWGDLIAHG 295
DB 218 IPILAVWFLSIDERFLNSKIHEAAKFTLPTMCLMLIYPLTFLAFGLGTFISOGLASG 277
QY 296 LQGLYDFGPGVGLLFGVLVSPITVITGLHQSPPTEL-ELENOGGSFIFATASMANIAQG 354
DB 278 YTFIYNLSPVAGAFNGAQWQVLVIFGIHWGFVPMINNLNLSRYGRDTMIAMVGPSNFAQA 337
QY 355 AACLAVFTAKSEKLGKLAGASVSVALGTEPAIFGVNLRWPFYIGITAAIGGALI 414
DB 338 GASLGVFLTKKPEVKAIAGSAALTOGFTGITEPSIYGVTLKYKPFVVIASIAAGAGAI 397
QY 415 ALPDIKAVAGAGAGFLGVVSDAPDW---MFLVCAVTFVIAFAAATAY-----GLYLV 466
DB 398 -----GAAGSSGAANA-IPGILTLPFGKGFVGFIL--GIAYAYILSAIGTYFF 444
QY 467 RRNGSIDPDATAAPVPAGTTKAEAPAEFSDNSTIIQAPLGEATALSVDAMFASOK 526
DB 445 --GYKDEKADGT---APTKEAKETGVE-----AEVIVSPIRGVIVPLNEVKDEAFSAGL 494
QY 527 LGSVAIVPTKGLVSPVSKIVVAPPSGHAPAVRTKAEDGNSVDILMHIGDFTVNLNGT 586
DB 495 LKGVALVPGEGKLIISFVNGTITAPTGHAGIIRS-----DRKVEILLHVGEDTVQLNGK 550
QY 587 HFNPLKKQSGDEVKAGELLCFFDIDAIAKAGYEVTTPIVSN 627
DB 551 YFKLLVAGQDRVLVGQALLFDELEAKADGYDITTPIVVTN 591

RESULT 13

Q9F8X3 PRELIMINARY; PRT; 480 AA.
ID Q9F8X3
AC Q9F8X3
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE TREP.
GN TREP.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17400;
RX MEDLINE=98027372; PubMed=9361421;
RA Gaballa A., Aboysinghe P.D., Ulrich G., Matthijs S., De Greve H.,
RA Cornelis P., Koedam N.;
RT "Trehalose induces antagonism towards Pythium debaryanum in
RT pseudomonas fluorescens ATCC 17400.";
RL Appl. Environ. Microbiol. 63:4340-4345(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17400;
RX MEDLINE=21030915; PubMed=11191810;
RA Matthijs S.L.C., Koedam N., Cornelis P., De Greve H.;
RT "The trehalose operon of Pseudomonas fluorescens ATCC 17400.";
RA Res. Microbiol. 151:845-851(2000).

DR EMBL; AF229829; AAC31031.1; -;
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR TIGRFAMS; TIGR00826; EIIB_glc; 1.
DR PROSITE; PS01035; PTS_EIIB_CYS; 1.
SQ SEQUENCE 480 AA; 51391 MW; 1FC73A88E302C013 CRC64;

Query Match

Best Local Similarity 39.4%; Pred.No. 5.8e-49;
Matches 186; Conservative 115; Mismatches 164; Indels 7; Gaps 5;

QY 2 DKDLAQRILRDIGEDNIVAAAHCASTRRLRLKDKTRQSDLDQDDDDDLKGFETGGMF 61
DB 4 DYSNIAREILLEHSGSDNLEQAAHCVTRLRALKDPSLVNASALNQVDLVKGSFFTGGLF 63
QY 62 QIIVGPGDVHVFKELDSDATSKDIAVSTEQKDVVANNANWFSRAVKVLADIFVLPIL 121
DB 64 QWVIGPGEVEKVVAAALREOTGL-AAATADYKKKGADKTNALQRLRVFSDVFMPIPAL 122
QY 122 VGGGLMAINLVVAQDLF-CPQSLVEMFPOISGVAEMINLMASAPFAFLPVLVGFATK 180
DB 123 IAGLLGINNLMGAKGMFIEGQILLLEAYPNLDGLWSLINLMANTSFFVLPALVGSAAK 182
QY 181 RFGNEFLGAGIGMAYFPTLVNGYD---VAATMTAGEMPMWSLFG-LDVAQAQYQTVL 236
DB 183 RFGSEILGVLGMLVHPDLNANWYKAVAGDGSPLPYDFIFGWFQIEKVYQGOIL 242
QY 237 PVLVSVTLATIEKFLHRLMGTDPLITPVLTLLTGLFTLFAIGPAMRWGDLIAHGL 296
DB 243 PILMAAYVSVTEKWLRAVPNAQLLVPTTIVTIGVLAALIGPVTRHLGILITEGV 302
QY 297 QGLYDFGPGVGLLFGVLVSPITVITGLHQSPPTELFN-QGGSFFATASMANIAQGA 355
DB 303 VTLFLAPWGAIFGLLYAPLVTGMHMFAYDLQLISTQGTFTWPMPTVMNLAQGS 362
QY 356 ACLAVFEFLAKSEKLGKLAGASVSVALGTEPAIFGVNLRWPFYIGITAAIGGALIA 415
DB 363 AALGVYMSRNARDKSMASSTSAISAYFGITEPAMEGVNLRKFPFYAALGSLGSIFLS 422
QY 416 LFDIKAVAGAGAGFLGVVSDAPDWMLVCAVTFVIAFAAATAYGLYLV 467
DB 423 LNKVQASAIQVGGPLPGF-SIIPIQYIPSEVIGVIAIVPVFVLTGCLSMKIVR 474

RESULT 14

Q8DNS8 PRELIMINARY; PRT; 627 AA.
ID Q8DNS8
AC Q8DNS8;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Phosphotransferase system enzyme II (EC 2.7.1.69).
GN SCRA OR SPRI566.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=1154234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenney M., McLeaster K., Mundy C.W., Nicot T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";

J. Bacteriol. 183:5709-5717(2001).
 RL EMBL: AF008524; RAL00370.1; -
 DR transferase; Complete proteome.
 KW SEQUENCE 627 AA; 65643 MW; ECA5A48877C03F46 CRC64;
 Query Match 28.6%; Score 956; DB 16; Length 627;
 Best Local Similarity 33.3%; Pred. No. 8.7e-49;
 Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;
 QY 1 MDKDLAQRILRDIGEDNIVAAHCAATRLRLVLDKTDVDRQSLDDDDPKLGTFTGGM 60
 DB 1 MNQEIATKVIDALGRENVSVAHCAATRLRVWVDEKINKEINLEKYGQAFNSGQ 60
 QY 61 FOIIVGPGVDHVEK-LLDQATSKDIAVSTEQKLDVVANNANWFSRAVKVLADIFVPL 117
 DB 61 YQIIFGTGVNMYDEVVVLGPTS-----SKDDMAEVAQOGNWFQRAIRTEGDFVFI 115
 QY 118 IPLVGGGLLMAINNVLVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAFLPVLGFT 177
 DB 116 IPVIVATGLFMGVGRGLFALEM-----PLPGDFATYQILTDFTAILPLGLVWVS 165
 QY 178 ATRKFGGFLGAGIGMAVFTLVNGYDVAAATMTAGEMPWMSFLGLDVAQAGQGVLP 237
 DB 166 TRVFGGNAVGIIVGLMGLVSGSLPNAWA--QGGEVATMNFPGF-IPVVGLOGSVLP 221
 QY 238 VLVVSMILATIEKFLHKLMTADFLITPVLTLLTGLFTTIAIGPAMRWGDLAHLQ 297
 DB 222 AFIIIVGAKFEKAVKVPVDIDLLVTPFVLLVNSILGLFVIGPVFVHVENYILATK 281
 QY 298 GLYDFGPGVGLLGLVYSPVITGLHQSFPEIELELFNQGGSFIF-ATASMANIAQGA 356
 DB 282 AILSMFPGGLGFLGVHQLIVVGVHHIFNLLEVLQAAADHANFNAILTAAMTAQGA 341
 QY 357 CLAVFLAKSEKULKLAGASGVAVLGITEPAIFGVNLRPLPFFVIGTGAIGGALLAL 416
 DB 342 TVAVGVKTKNPKLTKTAFPAALSFAFLGITEPAIFGVNLRPLPFFLSIAGAGGLASI 401
 QY 417 FDIKAVAGALGFLGVVSDADPMV-----MFLVCAVVTFFVIAFGAAIAYGLYVR 468
 DB 402 -----LGLAGTNGITIPGIMLVNGVQLPQYLLMVAVSFALGFALTVMFGY---- 449
 QY 469 NSGIDPDATAPVPAGTKAEAE-APAEFSNDSTIIQAPLTGEATALSVDAMFASGL 527
 DB 450 --EVEDATAAKQAQAEVAEKEEVAALONEILV--FTIVGDVVALADVNDPVSSEAM 505
 QY 528 GSGVAIYPTKGVLSPVSGRIIVAFPSGHAFVTRKAEDGSNVDILMHIGFTVNLNGTH 587
 DB 506 GQGIAPKPSQGVVVALADAEVSAFTPTGHAFLKTR----NGAEVLHIVGIDTVSMNGDG 561
 QY 588 FNPCLKGQDEKAGELCEFDIAIKAAQGVETPTIWSN---YKKTGPNVNYGLGEIEA 644
 DB 562 FEAKVAGCNKVGADVLGTEDSNKIRAAAGLDDTMTIVINTADYASVAPVAI--GSYSK 618
 QY 645 GANLLNV 651
 DB 619 GDVIEV 625
 RESULT 15
 Q8NZJ7 PRELIMINARY; PRT: 627 AA.
 AC Q8NZJ7
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Putative sucrose-specific PTS permease, enzyme II.
 GN SPY18.1881.
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=186103;
 RP SEQUENCE FROM N.A. [1]

Db 567 VAVGOAVKGDLLGHFDPKIAEAGLDDTTMMIVTNSADYQSVLILAGHVLIG 620

Search completed: October 1, 2003, 05:03:58
Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 05:00:51 ; Search time 27 Seconds
(without alignments)
1035.832 Million cell updates/sec

Title: US-10-019-284A-2
Perfect score: 3342
Sequence: 1 MDHKDLAQRILRDIGEDNI.....TEAGANLLNVAKEAVPATP 661

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.JA.*
1: /cgn2_5/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_5/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_5/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_5/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_5/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_5/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
1	923	27.6	628	4	US-09-107-532A-5288
2	826	24.7	631	4	US-09-107-532A-3902
3	737	22.1	583	4	US-09-107-532A-6811
4	644	19.3	590	4	US-09-107-532A-6004
5	528.5	15.8	243	4	US-09-134-001C-4114
6	511	15.3	478	4	US-09-134-001C-5065
7	495.5	14.8	427	4	US-09-107-532A-5530
8	437.5	13.1	482	4	US-09-107-532A-6644
9	401	12.0	527	4	US-09-634-238-368
10	329	9.8	696	4	US-09-134-001C-4296
11	315	9.4	656	4	US-09-107-532A-5083
12	296.5	8.9	585	4	US-09-107-532A-4366
13	271.5	8.1	470	4	US-09-107-532A-5798
14	267	8.0	170	4	US-09-134-001C-3134
15	230	6.9	686	4	US-09-252-991A-1815
16	183.5	5.5	98	4	US-08-936-165A-419
17	181.5	5.4	153	4	US-09-107-532A-4365
18	172.5	5.2	355	4	US-09-107-532A-5592
19	170.5	5.1	655	4	US-09-134-001C-5531
20	168.5	5.0	661	4	US-09-134-001C-4303
21	151.5	4.8	479	4	US-09-107-532A-6243
22	151.5	4.5	490	4	US-09-134-001C-5516
23	147	4.4	470	4	US-09-328-352-5397
24	146.5	4.4	387	4	US-09-107-532A-6002
25	144.5	4.3	498	4	US-09-107-532A-5037
26	142.5	4.3	451	4	US-09-107-532A-5352
27	142.5	4.3	487	4	US-09-107-532A-6319

28	142	4.2	495	4	US-09-252-991A-19278	Sequence 19278, A
29	142	4.2	575	4	US-09-328-352-7529	Sequence 7529, Ap
30	140	4.2	481	4	US-09-107-532A-6912	Sequence 6912, Ap
31	137.5	4.1	505	4	US-09-328-352-7470	Sequence 7470, Ap
32	135.5	4.1	359	4	US-09-134-001C-3728	Sequence 3728, Ap
33	134	4.0	353	4	US-09-134-001C-4325	Sequence 4325, Ap
34	134	4.0	651	4	US-09-252-991A-18065	Sequence 18065, A
35	133	4.0	403	4	US-09-328-352-7791	Sequence 7791, Ap
36	128	3.8	731	4	US-09-340-736E-1	Sequence 1, Appli
37	126.5	3.8	417	4	US-09-107-532A-5437	Sequence 5437, Ap
38	126.5	3.8	480	4	US-09-252-991A-17687	Sequence 17687, A
39	125.5	3.8	512	4	US-09-107-532A-5262	Sequence 5262, Ap
40	125.5	3.8	528	4	US-09-252-991A-32995	Sequence 32995, A
41	125	3.7	432	4	US-09-107-532A-6836	Sequence 6836, Ap
42	125	3.7	731	2	US-08-911-364-1	Sequence 1, Appli
43	125	3.7	733	3	US-08-464-700-2	Sequence 2, Appli
44	125	3.7	3519	3	US-09-428-517-4	Sequence 4, Appli
45	122.5	3.7	533	4	US-09-107-532A-6006	Sequence 6006, Ap

ALIGNMENTS

RESULT 1
US-09-107-532A-5288
; Sequence 5288, Application US/09107532A
; Patent No. 6383275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...628
; SEQUENCE DESCRIPTION: SEQ ID NO: 5288:
US-09-107-532A-5288


```

Query Match      27.6%; Score 923; DB 4; Length 628;
Best Local Similarity 32.3%; Pred. No. 3.9e-86;
Matches 215; Conservative 134; Mismatches 253; Indels 64; Gaps 12;

QY 1 MDKDLAQRILRDIGEDNIVAAAHCAIRLRLVKDKVDKQSLDDDDPDLKGFETGGM 60
DB 8 MDQAVGRVWEAVGGQKNVSLVHCATRLRFLKDKSLADTQKLKEDPDVQVQSGGQ 67
QY 61 FOIIVPGVDVHFKELDDATSKDIADVSTQELKDVVANNANWFRVAKVLIADIVPLPI 120
DB 68 YQWVIG-SNVADVYQISVD--EEGLAVS---EDGKEESKILNRLDIISSTFTPLGA 120
QY 121 LVGGGLMANNVLAQDLFGQSLVFNFPQISGVAEMINLWASAPAFPLVVLGVTATK 180
DB 121 MAAGVCLKFSL-----ATVWGWLTDSTGYQILFRAAGVTFPLVLAFTAAK 171
QY 181 RFGNEFLGAGIMAWPFTLVNGYDVAAATWAGEMPMSLFLGLDV--AQAGYQGTVLVP 238
DB 172 KFKANGLAVATAMALVYP-----AITAVAGAGEAISFFGIPVILSPSGYSSVPI 223
QY 239 LVVSWLATIEKFLHKLMTGADFLITPVLTLLTGLTFTAIQIPAMRWGDLIAHGLQ 298
DB 224 IIAVWQSKLEFPVKKVPOFLQMLVPLVVLVWVPLTFLALGPIGTAVNAGLGLFNS 283
QY 299 LXDGGPVGGLFLGLVYSPVITVTHQSPPIE-LELNNQGGSFIFATASVANTAQGAAC 357
DB 284 IYGFSPVAGLINGSWQVFNMGHGFVPMFLNIEQYGFVLMPLMLLPAILAQGAA 343
QY 358 LAVFELAKSEKILKLAGAGSVSAVLGITEPAIFGVNLRWLPFVYIGITAAIGALIALF 417
DB 344 LAVALRTKTKRALGISTVLSLGIETPTVYGVVTLPLKKPFTAAACISGGIGAILGS 403
QY 418 DIKAVA-----LGAAGFLGVYSIDAPDKMFLVCAVTVFVIAAGAAIAYGLVRRNGS 471
DB 404 GVKAFSSSLVLTPTFTINTVDGVSNTVAVIATGAFVLAFTVGLTILGFDQTDNQ 463
QY 472 IDPDATAAPVPAGTKAAEAAPAEFNSDSTIIQAPLTGEATLSSVSDAMFASGLASGV 531
DB 464 LE-----NKHANAGEPITSARHTLKSPLTKVLPSEVPDQVSSVGMGKI 510
QY 532 AIVPTKQLVSPVSGKIVVAFPSGHAFVTRKAEKGSNDVILMHGIFDVTNLNGTHNPL 591
DB 511 AIDPEVGLVAPADGEITFTFTGHAVGITT--TDGA--ELIHGMGTVESELNGFEIL 566
QY 592 KKGDEKAGELCEFDIDAIRAAQYEVPTPIVWSNXXKTGPVNTYGLGEIAGANLNV 651
DB 567 VKGDLVAKGDLIRFDIEAIRAAGVSVITPVVIN-----TDAFADILEL 612
QY 652 AKKEAV 657
DB 613 DQKEII 618

RESULT 2
US-09-107-532A-3902
; Sequence 3902, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Walcham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC

OPERATING SYSTEM: <unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3902:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...631
SEQUENCE DESCRIPTION: SEQ ID NO: 3902:
US-09-107-532A-3902
Query Match      24.7%; Score 826; DB 4; Length 631;
Best Local Similarity 30.5%; Pred. No. 4e-76;
Matches 196; Conservative 123; Mismatches 243; Indels 80; Gaps 11;

QY 1 MDKDLAQRILRDIGEDNIVAAAHCAIRLRLVKDKVDKQSLDDDDPDLKGFETGGM 60
DB 28 MDYSQLAKDIVRFVGGSENVSNVYHCATRLRFLTKDNKKADKEVQELGVITVVEAGGM 87
QY 61 FOIIVPGVDVHFKELDDATSKDIADVSTQELKDVVANNANWFRSA-----VKVLADIF 114
DB 88 FQVWVG-----NAVNEVIDVLSKQM-----KLEDDASSGKRGTEKKGLNSFIDMIAAVF 137
QY 115 VPLIPVLGGGLMANNVLAQDLFGQSLVFNFPQISGVAEMINLWASAPAFPLVPLV 174
DB 138 APTGLVLAGSLIKGVLAICTSLNLTTE-----SGTYILNAAADAFFYFLPIFL 188
QY 175 GFTATKRFNGNBFGLAGIGAMVFPVTVNGYDVAAATMTAGEMPMSLFGDLVQAQYQGT 234
DB 189 AYTAARKFNTDRFIAMVIAAALVPTIVSAYSISITL-----RELGMPIVLARYTST 240
QY 235 VLPVLVWSLTIATIEKFLHKLMTGADFLITPVLTLLTGLTFTAIQIPAMRWGDLIAH 294
DB 241 VIPAILAVWVLSYIEPKIRKSLHESIRNLITFTFCIIVMPVILLVWGPDIADIASQLIA 300
QY 295 GLOGLYDFGPGVGLFLGLVYSPVITVTHQSPPIEELFNQGGSFIFATASMANI-AQ 353
DB 301 GYLAVYNESPLVSGAVIGGFQWQVLYIFGLHWGLVPVMTNNLSFYGRDITLGPACMTAVAA 360
QY 354 GAACLAFFLAKSEKILKLAGAGSVSAVLGITEPAIFGVNLRWLPFVYIGITAAIGAL 413
DB 361 AGAVLGVFLTKNKKVKYLSLSAFITALEGITEPAVGYVTLKYKRPFFIACICGAIFGV 420
QY 414 IALFDIKAVALLGAAGFL-----GVWSIDAPDMVFMFLVCAVTVFVIAAGAAIAYGLYL 465
DB 421 AGAAGACALAVATRSILSPFIYIGEGFVWLVASFLAMISSCMLTFEGY----- 470
QY 466 VRRNGSIDPDATAAPVPAGTKAAEAAPAEFNSDSTIIQAPLTGEATLSSVSDAMFAS 525
DB 471 -----KDEIB---EESKDIVLSTPAAGEITLSEVNDTFTFASG 506
QY 526 KLGSGVAIVPTKQLVSPVSGKIVVAFPSGHAFVTRKAEKGSNDVILMHGIFDVTNLNG 581

```

RESULT 2

US-09-107-532A-3902
Sequence 3902: Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
Applicant: Lynn A Doucette-Stamm and David Bush
Title of Invention: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
Number of Sequences: 7310
Correspondence Address:
Addresser: GENOME THERAPEUTICS CORPORATION
Street: 100 Beaver Street
City: Waltham
State: Massachusetts
Country: USA
ZIP: 02354
Computer Readable Form:
Medium Type: CD/ROM ISO9660
Computer: PC

```

Db      507 SLGEGFAIITDCKIYSSVNGEYSTVFPTKHAIGV--VSEEGA--BILIHIGDITVNLNG 562
      586 THENPLKKGDEYKAGELCEPDIDAIAKAAGYEVVTPPIVYSN 627
      563 KYQSAVSDGKVKRKGDLLEMDQLBELIKEGYDPTTWIVTN 604

RESULT 3
US-09-107-532A-6811
; Sequence 6811, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6811:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...583
; SEQUENCE DESCRIPTION: SEQ ID NO: 6811:

US-09-107-532A-6811
Query Match      22.1%; Score 737; DB 4; Length 583;
Best Local Similarity 30.6%; Pred. No. 5.4e-67;
Matches 180; Conservative 119; Mismatches 239; Indels 50; Gaps 13;

Qy      87 VSTQLKDVVANNANWTSRAVKVLADIFVPLIPLVGGGLMAINNVLVA---QDLFGPQ 143
      7 VSTEGGAAAKQNLHPQRAIVLAIEFTPLIPAIIVGGLTLPNVLLEGIQFESLGG-- 64
      144 SLVMEFPOISGVAEMINLMASAPFAFLPVLVGETATKRFSGNEFLGAGIGMAMVFPVLN 203
      65 TIVSHKFWNGVNAFLWLPGBAIFHFLPVGTWTSIAKKMGTTQILGVLTIVSPOLLN 124
      204 GYDVAATMTAGEMPWMLFGDLDAQAGYQSTVPLVYVSWILATIERFLHKLRLMGTADEL 263

```

```

Db      125 AYSVAST-AAADIPFWDGFAQVQVMIGVQAQVIPAFLAGMLAYLEIFFRKYIPOSISMI 183
      264 TPVLTLLTGLTEIAIGP-----AMRWVGDLAHLGQLGDFGPGVGLFLGVSPIV 319
      184 FVPLFSLPTVLAHVILSPIGTWGWSINIVNTGLTSSIXW---LPSAVEGELYAPLV 240
      320 ITGLHQSFPPTLELFPNOGGSF-IFATASMANIAGGAACIAYFFLAK-SEKLKGLAGASG 377
      241 ITGLHMMNAIDMQLIADFGSTNLMFMIALNSIAQGSASVLAIIFLHRNKKKEGSLSPAM 300
      378 VSAVLGITEPAIFGVNLRMPFYIGIGTAAIGGALIALFDIKAVAGAACGFLGVVSDA 437
      301 ISCYLGVTPEMAGINLKYVYFVAAMVSGSLAGMFANLMGVRANAICVGGGLPGLAIOA 360
      438 PMVMFLCAVVTFVIAAGAAIAYGLYLVRRNG----SIDP--DATAAPVPACTTKAAEAA 492
      361 ETWVFPIIAMIIVAIPIEGLTIIF-----RRQGLNKIDIPAVPENAAOVQLOTANGATAT 415
      493 PAEFSNDSTIIQAPLTGAIA-----LSSVSDAMFASGLGSGVAIVPTK 537
      416 PQSPEPVS-----TGTAVATKETLFAVAAGTKEITEVNDPVFSQKMMGDGVAVEPSN 469
      538 GOLVSPVSGKIYVAPPSGHAFVARTKAEDGSNVDMILMHIGFTVNLNTHENPLKKQDGE 597
      470 GKVPYVNGKVTSPETKHAIGILS-----NEGLEVIVHMGDLTVELKGVPNFVFKGYL 525
      598 VKAGELLCFEDDAIKAAGYEVTTPIVYVSNKYKTKGPVNTYGLGTEAG 645
      526 VTPETLIAEMDLPEIEQAGKKTDIIVALTNNKXVAGLSLDQSGLVRPQ 573

RESULT 4
US-09-107-532A-6004
; Sequence 6004, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6004:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 590 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```
;
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...590
; SEQUENCE DESCRIPTION: SEQ ID NO: 6004:
US-09-107-532A-6004

Query Match      19.3%; Score 644; DB 4; Length 590;
Best Local Similarity 26.6%; Pred. No. 2.2e-57;
Matches 164; Conservative 130; Mismatches 249; Indels 74; Gaps 11;

QY 3 HKDLAQRILRDGGEDNIVAAAHCAIRLRLVLDKDVDRQSLDDPDLKQTFEFGGMFQ 62
DB 24 YQELARKIVENGGKENGINGLTHC:TRLRFLKNEEKANDEILKMGDGIIVVMRAGGQY 83
QY 63 IIVGPDVDHVKELDDA--TSKDIAVSTEQULKDVVANNANWFSRAVVLADIFVPLPI 120
DB 84 VVIG-NHVPVVEEIKAGNLTFDEAVTEKRP-----FMDLIDIGSCQFPFLAI 134
QY 121 LVGGGLMAINNVVAQDLFGPQSLVEMFPQISGVAEMINLMASAPFAPLVLVGFATK 180
DB 135 LAAGMIRGLTAFLV-----FLGAFDRGSGTFVMFDNIGDSVFQFMPVILGHTAAR 185
QY 181 REGNEFLGAGIGAMVPPTLVNGYDVAATMTAGEMPWMSLFGLDVAOA----- 229
DB 186 KKVNEFVGMLCAALMPSL----SLRALSAAEAPLTTFSGTIFEPAPIQTVFGIPW 241
QY 230 ---GYQGTVPLVNVSWILATIEKFLHKRLMGTADFLITPVLTLTLTGLFTFIATGPAMR 286
DB 242 IARNYASSVPIIFIVILLASQIQPKIKLVPEMIANFVFPEFTVLITMPLGFLLVGPVET 301
QY 287 WYGDLLAHGLOGLYDFGGVGLGLGLVSPVITVGLHOSFPPIELELFN-QGGSFIPAT 345
DB 302 FATDILMAGFELLALSPIYGAIVGFFWQIIVNFGLHWAIVPMGLMGSFVNGWGNIMTP 361
QY 346 ASMANIAQGAACLAFFFLAKSKLKLKLAGASGVSAVLGITEPAIFGVNLRKWPYVIG 405
DB 362 VAVVSGTAAALTYALFKLRNPKRAIPAIVSGIVGITEPAIYGFILPRKFIETCY 421
QY 406 TAAIGCALATDIDIKAVAGAGELGVVSI-----DAPDMVFLCAVAVTFVIAFGAIA 460
DB 422 GGAIGGAYSLMNLTSWNGGGIGITIPNVIKPDGLDIDVINVLIGIAIAMVVSITLFF 481
QY 461 YGLYLVRNGSIDPDATAAPVPAGTTKABEAPAEFSNDS--TIIQAPLTGSAIALSSVS 518
DB 482 F-----WKDEAGETDDIQKSGKEIVKTPIQGQIAPLNAAK 517
QY 519 DAMFASGLGSGVAIVPTKGQVSPVSGKIVAVFSGHAFVTRKAEDGNSVDILMHICF 578
DB 518 DAAFAOGTLGRGLIYPEKGEVRAPDGTVMTLFPKQAIMVSE----TGLELLIHVGL 573
QY 579 DTVNLNGTHFNPLKKOG 595
DB 574 DTVLEGKYFESLVQOG 590

RESULT 5
US-09-134-001C-4114
; Sequence 4114, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
```

```
;
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4114
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4114

Query Match      15.8%; Score 528.5; DB 4; Length 243;
Best Local Similarity 48.6%; Pred. No. 4.1e-46;
Matches 107; Conservative 43; Mismatches 55; Indels 15; Gaps 3;

QY 250 KFLHKRLMGTADFLITPVLTLTLTGLFTFIATIGPAMRWGDLAHLAQLOGLYDFGGPVGGL 309
DB 13 KGLHKVLTVDNLTLPLLSIFITAFITFLFVGPVTRQLGYWLSDLGLTWLYEFGGAIGGL 72
QY 310 LFLGVYSPVITVGLHOSFPPIELELF-----NQGSFIFATASMANIAQGAACLAFFFLAK 365
DB 73 IFGLLYAPIVITGMHHIFIAVETTLIADATKTGGSFIFPIATMSNIAQGAALAAFFIHK 132
QY 366 -SEKLKLAGASGVSAVLGITEPAIFGVNLRKWPYVIGITAAIGGALIALFDIKAVAL 424
DB 133 QNKKLKGVAAGASALLGITEPAMFVGNLKLRYPPFICAVAGSGIGNAYISFFKVAIAL 192
QY 425 GAAGFLGVWSIDAPD-----MYMFLVCAVAVTFVIA 454
DB 193 GTAGLPGFISINPTHAGLHLYLIGMLIAFVSVVTVLVS 232

RESULT 6
US-09-134-001C-5065
; Sequence 5065, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5065
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5065

Query Match      15.3%; Score 511; DB 4; Length 478;
Best Local Similarity 29.8%; Pred. No. 8.3e-44;
Matches 147; Conservative 90; Mismatches 199; Indels 58; Gaps 16;

QY 6 LAQRILRDIGGEDNIVAAAHCAIRLRLVLDKDVDRQSLDDPDLKQTFEFGGMFOIIV 65
DB 11 LAKDITHALGGSQNSIINIIHCHMTRVRIKVNDAKNVYDELKLSINGVLGVVE-DERIQVW 69
QY 66 GPGVDVDFVKELDDATSKDIAVSTEQLDKQVVA-----NNANWFSRAVK 108
DB 70 GPGVINKVAKLMADCSGATLAETETNQSYKSAEKRAYEHKKQFQSQKQSKW-NKVLK 128
QY 109 VLADIFVPLIPLVGV-----GGLLMAINNVVAQDLFGPQSLVEMFPQISGVAEMINLMAS 164
DB 129 SIANIFILIPAFIAGGLIGGIAILSNLLTAGSISG-----QWIIQIIVTVLVNITKD 180
QY 165 APFAPLVLVGFATKTRPGGNEFPLGAGIGAMWVPTLVNGYDVAATMTAGEMPWMSLFL 224
DB 181 GMLFYLAIFTGINSKAVEGATPGLGVIGGT-----TLTG-----ITDENPKNIFTG 229
QY 225 DVAOAGYOGTVLPVLVWSILATIEKFLHKRLMGTADFLITPVLTLTLTGLFTFIATGPA 284
DB 230 EHLAG-OGGIIGVIFAFWLLSNVKEKRLHKLIIPNSIDIIVTPTITLLIGLITFIIMPL 288
```



```
QY 311 FGLVSPVITGLHQSF-----PPIELEFNQ-----GSFIF 343
| | | | |
DB 237 FGIKRLIFGLHLHFHAFWPFSGSKNAAGEIRIGDQRIEIREGAHUTSGFMQ 296
| | | | |
QY 344 ATASMANIAGAACLAFFLAKSEKLAG-----ASGVSALVIGITEPAIFGVNLRWP 399
| | | | |
DB 297 GEFFVMFGLPAALAIYOTAKENKVVAGLMISAALTSLTGTITEPLEF--SFLFVAP 354
| | | | |
QY 400 FYIGIGTAAIGGALIALFDIKAVAG--AAGFLGVVSD-APDMVFLVCAVTVFIAP 455
| | | | |
DB 355 FLVIRAVLDGLSFLY--LNLVHLGYTSGGFDIVLLGIDLPNKTAWMLVIPVGIYAV 413
| | | | |
QY 456 GAAIAXGLYLVRN-----OSIDPDA-----TAAPVPAGTTKA-----488
| | | | |
DB 414 IYFVFRFLIVKENYKTPGREDKSSVTTTSASQLPFDVLKAMGGRENKHLHACITRLR 473
| | | | |
QY 489 -----EAFAPAEFSN 498
| | | | |
DB 474 IQNEKSKVDVAGLKLGLSGVLEVNNNQAIIFGPKSDQIKHDMAKIISGEITKPSSETI 533
| | | | |
QY 499 D-----STIIQAPLTGEAIALSSYSDAMFASGKLGSGVAIVPTKGQVSPV 544
| | | | |
DB 534 DEVSDESHVEDIVETIYAPGHGIIPLSEVPDKVFSEKMGMDGIGFVPSDKIVAPF 593
| | | | |
QY 545 SGRIVVAEPGSHAFVTRKAEDGSNVDILMHIGFDVNLNGTHFNPLKKGDEYKAGELL 604
| | | | |
DB 594 DGTVKIETPKHAIGL-----ESDSGVEVLHIGIDIVKLNGBESLVNTDEPVTGGQPL 649
| | | | |
QY 605 CEFDIDAKAAGVETVTPVVSNYK-KTGVPNTYGLGEIEAGANLLNV 651
| | | | |
DB 650 MKIDLEVLEHAPSIIPTVIITNQEDKT--LTIEDVKQVDPGKAIMTI 695
| | | | |

RESULT 11
US-09-107-532A-5083
; Sequence 5083, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arianello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5083:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 656 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1....656
; SEQUENCE DESCRIPTION: SEQ ID NO: 5083:
US-09-107-532A-5083

Query Match
Best Local Similarity 9.4%; Score 315; DB 4; Length 656;
Best Local Similarity 22.2%; Pred. No. 2.5e-23;
Matches 151; Conservative 96; Mismatches 207; Indels 226; Gaps 27;

QY 118 IPILVGGGLLMAIN-----NVLVAQDL-----FGPQSLVEMFPQIS 153
| | | | |
DB 59 LPLLFAVLGLGMSVDKSGAAALAGLVAFEPVNVYKTDTSVATLLNINIKPESVDSFQAIS 118
| | | | |
QY 154 GYAEMINLMASAPFAFLPVLVGYFTATKFRGGNEFLGAGIGMAMVFPPTLVNGYDVAATMTA 213
| | | | |
DB 119 NV-----FIGIAGLIAAALY--NRFEAKLPALSF-----148
| | | | |
QY 214 GEMPWNSFLGLDVAAGAQGVPLVPLVVSWSILATIEKFLHRLMGTAFLITPVTLLLT 273
| | | | |
DB 149 -----FSGKRLVPIAFAFVNLISAYL-----LLVMPFVSALV 182
| | | | |
QY 274 GFLTEIAGPAMRWYGDLLAHGLQGLYDFGGPVGGLFLGLVSPVITGLHQSPPI---330
| | | | |
DB 183 SFAKFI--SLGMWG-----AGLYGF-----FNRLIIP--TGLEQAMNSVFWF 221
| | | | |
QY 331 ELELENQSGSFIFA--TASMANIAGQ-----AACLAVFFLAKSE-KLKG---LA 373
| | | | |
DB 222 DVAGINDIGKFLASEGTGVTGMVQAGFFPIMMFLGPAGAYAIYRNARPENKAKTASLML 281
| | | | |
QY 374 GASGVSAVLGITEPAIFGVNLRWRPFYI--GIGTAAIGGALIALFDIKAVAGAA--428
| | | | |
DB 282 AAFAFSFTGTVEPLEFSF-MFVAPLVYLHAIET-GISLAVSAFFHWTAGFAFSAGFVD 339
| | | | |
QY 429 -FLGVVSDADPMVNFVCAVVTVI-AFG-----456
| | | | |
DB 340 YFLSLKNPIANOPLMLIVQGLVTAIVYVYGFNFIAKFRFLHMTPGREEDLSDDDTATTNS 399
| | | | |
QY 457 ----AAIAYGLV-----VRRNGSIDPD--ATAAP-----480
| | | | |
DB 400 DNKIAAQAASKIYAAALGGADNVTSDNCTRLRLQVKDTGLIDQNKIKATGVPGMKIIDGX 459
| | | | |
QY 481 ---VPAGT-----TKAEAPAEFSNDSTIIQ-----APLTGEAIAL 514
| | | | |
DB 460 NAOQVIVGTVEQFVADEMAKLHGGAAARPAQTNTVVKTFAGETIEQDIAIANGKLIPI 519
| | | | |
QY 515 SSVSDAMPFASGKLGSGVAIVPTKGOLVSPVSGKIVVAFPPSGHAFVATKAEDGSNDVILM 574
| | | | |
DB 520 TEVSDDVFAEKMMGDGYAVLFENGEIFSPIAGTTNIPPTKHAVGICQIDA----GIEVLL 575
| | | | |
QY 575 HIGFDTVNLNGTHFNPLKKGDEYKAGELCEFDIDAIIKAAGYEVTTPIVVSNTYKKTGPV 634
| | | | |
DB 576 HMGINTVDLKGEPFTLYVEEGQKVARGOLIALVDLAAIQSAGKNTDMVWVFTNGDKVQSL 635
| | | | |
QY 635 NTYGLGEIEAGANLLNVAKK 654
| | | | |
DB 636 EIEPARDVKANDKIGSVSNK 655
| | | | |
```

RESULT 12

```
US-09-107-532A-4366
; Sequence 4366, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
```


Query Match 8.1%; Score 271.5; DB 4; Length 470;
Best Local Similarity 21.1%; Pred. No. 4.3e-19;
Matches 116; Conservative 78; Mismatches 143; Indels 213; Gaps 21;

QY 208 AATWAGEMPWISFGLDVAQAGYQGVLPVLVVSUWLAIEKFLHRLMGADFLTPV 267
DB 14 AGTQVGDPLWLAWATDLNLKAGAD-----MSQYEFVLHWT----- 52

QY 268 LTLITGFLTEAIGPAMRWYGDLLAHGLOGLYDFGGVGLFLGLVSPVITGLHQS 327
DB 53 -----PAREKVGQMGIS-----SGILMGAF-----MYRNV 79

QY 328 PPIELFNOGGSIFATASMANIAQGAACLAFLAKSEKLGKLAGASGVSAYLGTPE 387
DB 80 DPKRKARYK-----SMYF-----SAALAV-----LTGVTEP 106

QY 388 AIFGVNLRLPFFYIGICTAIGGALIALFEDIKAVAGAAAGFLGV-----VSIDA---PD 439
DB 107 LEF-MEMFAAFLAVIY--SVIQGAFAFAMADILPLRVESFGNIELLRTPLAKAGLGGD 163

QY 440 MYMLVCVAVTVFAAGAAATAYGL--YLV-----RNGSIDPDATAAPV-AGTKA 488
DB 164 LINEVLC-----VIAFG-VVTYFLANFLIKFNFATPCRNNGYDNDSEETVVSNSGTGA 217

QY 489 EAE-----APA 494

DB 218 DQVVQIHLGGKENIKVDYDACHTRLRVSVNDREKVGTEAMKRAGAMGLIVKDNQVQA 277

QY 495 EFSNDSIIQAPL-----TGEAIALS 515

DB 278 VYGPADVLKSDIEDLLQSGAEIPEPEMVSTGNGQAGSKOYLTEQELVSAASGEVTPLS 337

QY 516 SVSDAMFASGLGSGVAIVPTKGQVSPVSGKIWVAPPSGHAFVATKAEAGSNVDILMH 575
DB 338 EVKDPVFSQKMGDGFVITPSREVAPVIAKVTISFPSKHAIGLETK--DG--IEVLII 393

QY 576 IGFTVNLNGTHFNPLKKQGEVKGAGELCEFFDIDAIAKAAGYEVTPPIVVSNNKKTGPVN 635
DB 394 MGIDTVQMNOPAFETIAVKEGQEVGTGKLAEMDLVDVQNEEKDPTIMIVFTN-DKVEEV 452

QY 636 TYGLGEIETAG 645

DB 453 IKQLGTWTAG 462

RESULT 14
US-09-134-001C-3134
; Sequence 3134, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3134
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3134

Query Match 8.0%; Score 267; DB 4; Length 170;
Best Local Similarity 39.8%; Pred. No. 2.2e-19;
Matches 53; Conservative 31; Mismatches 45; Indels 4; Gaps 2;

QY 503 IQAPLTGEAIALSVSDAMFASGLGSGVAIVPTKGQVSPVSGKIWVAPPSGHAFVAT 562

DB 22 IYAPLTGEYVKIEDIDPVPFAQKMGEGGGINTEGEVSPIEGKVDNVFPTKH--AVGL 79
QY 563 KADGSGNVILMHIGDFTVNLGTHFNPLKKQGEVKGAGELCEFFDIDAIAKAAGYEVTP 622
DB 80 KAENG--LELLVHIGDFTVQLDQKGFVLSVESGDDIKIGDPLIRFDLEYINNNKSIISP 137
QY 623 IVVSNNKKTGPVN 635
DB 138 IITNSDQTESIH 150

RESULT 15
US-09-252-991A-18115
; Sequence 18115, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18115
; LENGTH: 686
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18115

Query Match 6.9%; Score 230; DB 4; Length 686;
Best Local Similarity 32.2%; Pred. No. 1.6e-14;
Matches 59; Conservative 40; Mismatches 68; Indels 16; Gaps 6;

QY 488 ABAEAPAE-----FNSNDSTIIQAPLTGEAIALSVSDAMFASGLGSGVAIVPTKGQVLS 542
DB 1 AKSPETEPEPAPMNNKNLAKPLSGVPVPLNRVPDPVFSSTGLGEGIAIDPLNDCLHA 60

QY 543 PVSGKIWVAPPSGHAFVATKAEAGSNVDILMHIGDFTVNLGTHFNPLKKQGEVKGAG 602
DB 61 PCAGLVSHLARTRHLSLR--ADNGA--ELLHVGLDVTQLQGEFGEALVEEGARVIEGQ 116

QY 603 LLCEEDIDAIAKAAGYEVTPPIVVSNNK--KTGPNVTYGLGEIETAGANLLNVA--KKEAVP 658
DB 117 PLLRFDLRVARGSRSLITVMILTINGDGFQVRPLTT---NPVEVGAPLLQUSPEKAEQRP 173

QY 659 ATP 661
DB 174 ANP 176

Search completed: October 1, 2003, 05:05:41
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2003, 05:04:06 ; Search time 217 Seconds
(without alignments)
481.930 Million cell updates/sec

Title: US-10-019-284A-2
Perfect score: 3342
Sequence: 1 MDKDLAQRILRDIGGEDNI.....IEAGANLLNVAKEAVPATP 661

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues 587654
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications_AA:*
- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep:*
 - 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep:*
 - 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep:*
 - 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep:*
 - 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep:*
 - 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep:*
 - 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep:*
 - 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep:*
 - 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep:*
 - 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep:*
 - 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep:*
 - 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3310	99.0	661	10	US-09-738-626-6404
2	956	28.6	627	9	US-09-815-242-13467
3	725.5	21.7	683	10	US-09-738-626-6961
4	550	16.5	484	9	US-09-815-242-10809
5	538	16.1	455	9	US-09-815-242-13794
6	522	15.6	484	9	US-09-815-242-12272
7	490.5	14.7	439	9	US-09-815-242-5660
8	472	14.1	545	15	US-10-156-761-11459
9	438	13.1	474	9	US-09-815-242-10243
10	368.5	11.0	679	9	US-09-815-242-5658
11	368.5	11.0	679	9	US-09-815-242-12270
12	321	9.6	719	9	US-09-815-242-12942
13	319	9.5	687	9	US-09-815-242-5822
14	308.5	9.2	648	9	US-09-741-669-330
15	308.5	9.2	648	9	US-09-815-242-10089

16	290.5	8.7	551	9	US-09-815-242-10500
17	233.5	7.0	640	9	US-09-815-242-10786
18	232	6.9	280	9	US-09-815-242-12479
19	226.5	6.8	842	9	US-09-815-242-11950
20	221.5	6.6	263	9	US-09-815-242-13018
21	221.5	6.6	263	9	US-09-815-242-13152
22	220.5	6.6	135	9	US-09-815-242-5842
23	209	6.3	149	15	US-10-156-761-14503
24	201.5	6.0	142	9	US-09-815-242-5718
25	187	5.6	557	9	US-09-815-242-12301
26	183.5	5.5	98	9	US-09-939-980-419
27	182	5.4	643	9	US-09-815-242-5394
28	173.5	5.2	650	9	US-09-815-242-13341
29	173.5	5.2	650	11	US-09-769-787-95
30	173	5.2	704	15	US-10-156-761-11225
31	172.5	5.2	658	9	US-09-912-020-352
32	168.5	5.0	422	15	US-10-156-761-12701
33	167	5.0	639	12	US-10-238-075-607
34	160.5	4.8	448	11	US-09-769-787-56
35	157.5	4.7	482	9	US-09-815-242-5401
36	157.5	4.7	488	9	US-09-815-242-12564
37	156	4.7	585	9	US-09-815-242-11923
38	150.5	4.5	556	9	US-09-815-242-11023
39	141.5	4.2	454	9	US-09-815-242-10731
40	140.5	4.2	510	9	US-09-815-242-13474
41	140	4.2	534	15	US-10-156-761-12074
42	138.5	4.1	505	11	US-09-769-787-120
43	138.5	4.1	510	9	US-09-815-242-13607
44	138	4.1	475	9	US-09-815-242-13774
45	138	4.1	502	15	US-10-156-761-8917

ALIGNMENTS

RESULT 1

US-09-738-626-6404
; Sequence 6404, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6404
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6404

Query Match 99.0%; Score 3310; DB 10; Length 661;
Best Local Similarity 98.9%; Pred. No. 1e-290;
Matches 654; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 MDKDLAQRILRDIGGEDNIIVAAAHCAATRLRLVLKDTKDVRQSLDDDPDLKGTFTGGM 60

Db 1 MDKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLKDTKDVDRQSLDDPDLKGTFTETGGM 60
QY 61 FOIIVPGDGVHVFELDDATSKDIADVSTEQDKVYVANNANWFSRAVKVLADIFVPLPI 120
Db 61 FOIIVPGDGVHVFELDDATSKDIADVSTEQDKVYVANNANWFSRAVKVLADIFVPLPI 120
QY 121 LVGGGLLMAINVLVAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTATK 180
Db 121 LVGGGLLMAINVLVAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGTATK 180
QY 181 REGGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLPVLV 240
Db 181 REGGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLPVLV 240
QY 241 VSWILATIEKFLHKLRLMGTADELITPVLTLLTGLTFTTAIGPAMRWVGDLLAHGLQGLY 300
Db 241 VSWILATIEKFLHKLRLMGTADELITPVLTLLTGLTFTTAIGPAMRWVGDLLAHGLQGLY 300
QY 301 DFGGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIFATASMANTAQGAACLAY 360
Db 301 DFGGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIFATASMANTAQGAACLAY 360
QY 361 FFLAKSEKILKLAGASGSVAIGITEPAIFGVNLRNRPFFYIGTAAIGGALIALFDIK 420
Db 361 FFLAKSEKILKLAGASGSVAIGITEPAIFGVNLRNRPFFYIGTAAIGGALIALFDIK 420
QY 421 AVALGAAGFLGVVYSIDAPDMWFLCAVVTETAFGAALAYGLYLVRNGSIDPDATAAP 480
Db 421 AVALGAAGFLGVVYSIDAPDMWFLCAVVTETAFGAALAYGLYLVRNGSIDPDATAAP 480
QY 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATLSSVSDAMFASGLSGVAIVPTKQGL 540
Db 481 VPAGTTKAEAPAEPSNDSTIIQAPLTGEATLSSVSDAMFASGLSGVAIVPTKQGL 540
QY 541 VSPVSKIVVAPSPGHAFVRKAEDGSNVDMILMHIGDPTVNLNGTHFNPLKKQGDVKA 600
Db 541 VSPVSKIVVAPSPGHAFVRKAEDGSNVDMILMHIGDPTVNLNGTHFNPLKKQGDVKA 600
QY 601 GELLCEFFDIDAKAAGYEVTTPVSNYKKTGPVNTYGLGETEAGANLNLVAKKEAVPAT 660
Db 601 GELLCEFFDIDAKAAGYEVTTPVSNYKKTGPVNTYGLGETEAGANLNLVAKKEAVPAT 660
QY 661 P 661
Db 661 P 661

RESULT 2

US-09-815-242-13467
; Sequence 13467, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13467
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13467

Query Match 28.6%; Score 956; DB 9; Length 627;
Best Local Similarity 33.3%; Pred. No. 1.2e-77;
Matches 222; Conservative 141; Mismatches 246; Indels 58; Gaps 14;

QY 1 MDKDLAQRILRDIGGEDNIVAAAHCAATRLRLVLKDTKDVDRQSLDDPDLKGTFTETGGM 60
Db 1 MNQEIAKKVIDALGGRENVSVAHCAATRLRVVVKDEKINKEVLENKVKQGAFFNSGQ 60
QY 61 FOIIVPGDGVHVFKE---LDDATSKDIADVSTEQDKVYVANNANWFSRAVKVLADIFVPL 117
Db 61 YQIIFGTGTVNMKYDEVVVLGLPTS-----SKDDMKAEVAKQGNWFQRAIRTEGDFVFI 115
QY 118 IPIVLGGGLLMAINVLVAQDLFGQSLVEMFPQISGVAEMINLMASAPFAFLPVLVGT 177
Db 116 IPVIVATGLFVGSRGLFNALEM-----PLPGDFATVTOILTDTAFIILPGLVWS 165
QY 178 ATKREGNEFLGAGIGMAWVFTLVNGYDVAATMTAGEMPWMSLFGLDVAQAGTGTVLP 237
Db 166 TFRVGGNPAGVYGLGKMLVSGSLPNAWAVA---OGGEVTAMNFFGF-IPVVGQGSVLP 221
QY 238 VLVYSWILATIEKFLHKLRLMGTADELITPVLTLLTGLTFTTAIGPAMRWVGDLLAHGLQ 297
Db 222 AFILGVGAKFEKAVRVVDPVLDLLVTPFTVLVMSILGLFVIGPVHVVENYLIATK 281
QY 298 GLYDFGPGVGLLGLVYSPVITGLHQSFPPIELFNQGGSFIF-ATASMANIAOGAA 356
Db 282 AILSMFGLGGLGVLGGVHQLIVSGVHHIFNLLEVQLLAADHANFPFNAITAAMTAOGAA 341
QY 357 CLAVFFLAKSEKILKLAGASGSVAIGITEPAIFGVNLRNRPFFYIGTAAIGGALIAL 416
Db 342 TVAVGVTKNPKLKTAFPAALSAPLGTETPAIFGVNLRNRPKPFELSIAAGIAGGLASI 401
QY 417 FDIKAVALLGAAGFLGVVYSIDAPDMV-----MFLCAVVTETAFGAALAYGLYLVR 468
Db 402 -----LGLAGTNGITIIPTGMLYVNGQLPYLLMVAVSFALGALTMYFGY----- 449
QY 469 NGSIDPDATAAPVPAGTTKAEAE-APAEFSNDSTIIQAPLTGEATLSSVSDAMFASGKL 527
Db 450 --EVEDATAAKQAEVAEKEEVAAPALQNETLV--TPIVGDVVALDVNDPVFSSGAM 505
QY 528 GSGVAIVPTKQGLVSPVSGKIVVAFPPGHAFATKADGSDNVDILMHIGDPTVNLNGTH 587
Db 506 GOGIAVKPSQGVVVALADAESAISIAFPTGHAFGLKTR-----NGAEVLIHVGIDTYSMAGDG 561
QY 588 ENPLKKQGDVKAAGELCEFFDIDAKAAGYEVTTPVSNYKKTGPVNTYGLGETEAG 644
Db 562 FEAKVAQGNKVKAGDVLGTGTFDSNKTAAAGLDDTTWIVTADYASVAPVAT---GSYSK 618
QY 645 GANLNV 651
Db 619 GDAVIEV 625

RESULT 3

US-09-738-636-6961
; Sequence 6961, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

```

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6961
; LENGTH: 683
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6961

Query Match          21.7%; Score 725.5; DB 10; Length 683;
Best Local Similarity 30.9%; Pred. No. 9.9e-57;
Matches 216; Conservative 119; Mismatches 286; Indels 79; Gaps 18;

QY 7 AQRILRIGGDNVAAAHCAATRLRLVLTQKDVDRQSLDDDDPKLGTFTGTG--GMFQII 64
DB 9 SOHLENLGGPDNITSMTHCATRLRFQVQKQSIYDQDEIDSPVLGVVPGSGTGM-QVV 67
QY 65 VPGDGVHVFEL-----DDATSKDIAYSTOLKDVYANNANWFSRAVKVLADI 113
DB 68 MG-GSVANYQYQIILKDKMKHFDGEATES----SSKXGKGVGRKYSWIDYAEFLSDT 122
QY 114 FVPLIPVLVGGGLMAINNVLVAQDLFGPQSLVEMFPOISGVAEMINLMASAPFAFLPVL 173
DB 123 FRPILWALLGASLIITL--LVLADTFGLQDFRAPDEQPTTYVFLHSMRVSRYFLPIM 179
QY 174 VGFTATKRFNGNEFLGAGIGMAMVFPPLVNGYDVAATWAGEMPWMSLFLGLDVAQAQYG 233
DB 180 VGATAARLGNANIGWGAIPAALLTP-----EFLGASAGDTV--TVFGLPMVLYNDYSG 231
QY 234 TVLPVLVYSWTILATIEKFLKRLMGTADELITPVTLLTGLFLFAIGPAMRWVGDLLA 293
DB 232 QVFPPLIAAGLYWVEKGLKLIPEAVQMVFVPPFSLIMIPATAFLLGPPGIGVNGIS 291
QY 294 HGLQGLYDFGPGVGGLLFLGLYSPITVITGLHOSPPPIELELFNO--GGSFIFATASMANIA 352
DB 292 NLLEAINNFSFILLSIVIPFLYPLVPLGLHPLNAINIQNINTLGYDFIQGPMGAWNFA 351
QY 353 QGACLAFFFLAKSKLGLAGAS--GYSAVL--GITEPATFGYNLRWRPFYIGTAA 408
DB 352 CEGLVTVGFLLSIKERNAMQVSLUGGLAGLLGGISEPSYGLLRFLKTYIFLLPGCL 411
QY 409 IGGALIALFDIKAVAGLGAAGFLGVVSDIDAPDMVFLVCVVTFVIAF--GAIAIYGLXLVR 467
DB 412 AGGIVMGIFDIKAYAFVFTSLTIPAMD--PWLGYTIGIAVAFVYSMFLVALDY-----R 465
QY 468 RNSGTD-----PDAT-AAPV-----PAGTTKAEAEAPA 494
DB 466 SNEERDERAKVAADQAEEDLKAEANATPAAPVAAGAGAGAGAGAAGAAATAVAAPK 525
QY 495 EFSNDSSTIIQAPLTGEAIALSVSDAMFASGKLGSGVAIVPTKGLVSPVSGKIVAPPS 554
DB 526 LAAGEWVDIVSPLECKAIPLSVPDPPIFAAGKLGPGIALQPTGNTVAVAPADATVLYQKS 585
QY 555 GHAFVATKAEKSGNSVDILMHIGFDTVNLNGTHFNPLKQKQDEVKAGELLCEFFDIDA 614
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6961
; LENGTH: 683
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6961

Query Match          16.5%; Score 550; DB 9; Length 484;
Best Local Similarity 29.7%; Pred. No. 4.5e-41;
Matches 149; Conservative 99; Mismatches 180; Indels 74; Gaps 16;

QY 4 KDLAQRILRIGGDNVAAAHCAATRLRLVLTQKDVDRQSLDDDDPKLGTFTGTGFMFQI 63
DB 7 KRIADGIIYKEVGQENVDKVIHCHMTRVMDIRDYDKVDIEGLKKIDGVMGVVEDDTL-QV 65
QY 64 IVPGDGVHVFELKDD-----ATPSKDIAYSTOLKDVW-----A 97
DB 66 VVPGGVTKVQAQWVDQAGVKLGEFPFHGTTD--ASAGKSGKDLVEEKAQKAKQ 124
QY 98 NNANWFSRAVKVLADIYFPLIPILVG-----GGLLMANNVLVAQDLFGP--QSLVEMFPQI 152
DB 125 NNTSPKKVLKATISSIFVPMIPAFVAGIIGIAGVMSNLVWAGDISASWQYI----- 178
QY 153 SGVAEMINLMASAPFAFLPVLVGTFTATKRFNGNEFLGAGIGMAMVFPPLVNGYDVAATMT 212
DB 179 ----DVNLIITKIGIFAYLALYTGINASVFCATPALCGVIG-----AVTML 220
QY 213 AG----EMPWMSLFLGLDVAQAQYQGTGTVLPVLVSVNLIATIEKFLHKLMTADFLIIPVLT 269
DB 221 TGMNPDAPISNIFTGTTLSAG--QGGIIGVIFAVALLSLEKQLHKIVPESIDIIVTPTIS 279
QY 270 LLLTGLTFTFAIGPAMRWVGDLLAHGLQGLYDF-----GGPVGGLLFLGLVYSPVITGLHQ 325
```

```
Db 280 LLVIGLATIELIMP-----VAGAINSLGVGINVLEKGGVAGCTGLTGLTLPWMEGLHQ 335
QY 326 SFPIEIELENOGS-SPIFATASMANIAOAGACIAYFFELAKSEK--LKGLAGASGYSAVL 382
Db 336 ILPIHIEMINQMGTLULFILMAGAGQVGAALWIRCKSKVLVEMIKGALPV-GIL 394
QY 383 GITEPAIFGVNLRWPFYIGIGTAAIGGALIALF-DIKAVALGAAGFLGVWSIDAPDMV 441
Db 395 GIGEPILYGVTLPLGRPEITACIGGGIGGAGVAFAGFNGVGAIGPSSVALIPLIANNQWL 454
QY 442 MFLVCAVTVFIAPGAIAAYGL 463
Db 455 AYVLGLLAAYAGGEVATLFFGI 476

RESULT 5
US-09-815-242-13794
; Sequence 13794, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13794
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Salmonella typhi
; NAME/KEY: VARIANT
; LOCATION: (1)...(455)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13794

Query Match 16.1%; Score 538; DB 9; Length 455;
Best Local Similarity 30.7%; Pred. No. 5e-40;
Matches 141; Conservative 91; Mismatches 192; Indels 36; Gaps 10;

QY 6 LAQRIILRIGGEDNIVAAAHCAATRLRLVLKTDKVDQSLDDDDPDLKGTETGGMFOIIV 65
Db 9 LASDILRIGGEQNILLENLNCMTVRVVEQDDSLDIPRLKALPGVSGYVKQGEHQGLIV 68
QY 66 GPGVDVHVEKELDDATSKDIAVSTEQCLKDVYVANNANW-----PSRAVKVLADIFVPL 117
Db 69 GPGKAAQV----DAMRVQIRAGVKFDDAMARTKSEAKYKAPMSDALUKLANFIPL 124
QY 118 IPIVLGGGLLMAINNVLAODLFGPQSLVEMFFPQISGVAEMINLMASAPFAFLVVLGFT 177
```

```
Db 125 IFAFIASGLITGIINILKRPDIVG--DVAVHYPNLLG---LMGIFGSVAFSIMNIVGVN 179
QY 178 ATKRGGEHEFCAGIGMAWFFTLVNGVDVAATMTAGEMPWMSLFGDLVQAQGYOCTVLP 237
Db 180 TAKVFGGSQALG-GV-----MAGILSSPOLAQITLFG-EALOPG-RGGVIA 222
QY 238 VLVVSWILATIEKFLHKLRLMGTADELITPVLTLTLLTFTAIQGPAMRWGDLHLAHGIQ 297
Db 223 VLLVVALMCWIERQFRLKPLGSLLEILNPLLTVTITGAVAIVALQPLGWNISDAIHGAS 282
QY 298 GLYDFGPPVGGLLFGLVYSPIVITGHOSFPPIELEFN-QGGSFIFATASMANIAQAA 356
Db 283 WAIDRGGFLVGAVLAGTFLPLVLGLHGOGLVXHVLYQAHXYNALFFILAMAGYQIGA 342
QY 357 CLAVFFLAKSEKRLKGLAGASGYSAVLIGITEPAIFGVNLRWPFYIGITGAIGGALIAL 416
Db 343 AIAVLMKTRNARLKKVIRKALPVGLLIGICEPLIFGVTLPLGKPFPGIACLGAGVAGGALLISY 402
QY 417 FDIKAV-ALGAAGFLGVWSIDAPDMVFLVCAVTVFTIAF 455
Db 403 WKVATVITFGISGLPLALTIVAGKVLDFLLGYLIAIAGF 442

RESULT 6
US-09-815-242-12272
; Sequence 12272, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12272
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12272

Query Match 15.6%; Score 522; DB 9; Length 484;
Best Local Similarity 29.3%; Pred. No. 1.6e-38;
Matches 155; Conservative 84; Mismatches 188; Indels 102; Gaps 17;

QY 4 KDLAQRIILRIGGEDNIVAAAHCAATRLRLVLKTDKVDQSLDDDDPDLKGTETGGMFOI 63
Db 5 QQLAERLITAAVGSMDNIDSVNMCMTVRVIRKVLIDENKNVDQDLRHIDGVNGVIH-DERIQV 63
QY 64 IVGPGDGVHVFKELDATSKDIAVSTEQCLKDVYVANNANW----- 101
```

Db 64 VVGPRTVKNVANHMAE-----LSGVKLGDTIPHNHNDSEKMDYKSYAADKAKANKEAH 116
QY 102 -----WFSRAVKVLADIPEVLPIILVG-----GGLLMAINNVVAQDLFGPOSLEMFPP 150
Db 117 KAKQNKGLNKVLKSIANIFILPIPAFIGAGLIGGIAAUNSLNLMVAGYISG-----A 168
QY 151 QISGVAEMINLMASAPFAELPVLVGFTATKRFSGNEFLGAGIGMAMVFPPLVNGYDVAAT 210
Db 169 WITQLITVFNVIKDMGLAYLAIFTGINAAKEFGATPGLGSGVIGGT-----TLLTG----- 218
QY 211 MTAGEMPWMSLGLDVAQAQYOGTVLPVLSVWILATIEKFLHKLMTADFLITPVLTL 270
Db 219 -IAGKNILMNVTGEPLQPG-QGGIIGVIFAVWILSIVERLKHIVPNAIDIIVTPTIAL 276
QY 271 LLTGFLTFIATGPAMRWGDLAHLGLOGLYDFGPGVGLLGLVYSPIVITGLHQSPPI 330
Db 277 LIVGLLTIETIEMPLAGFVSDLSVSVNGIISIGVFSGFIIGASFLPLVLMGLHHIFTPI 336
QY 331 ELEFNQ-QGSFIFATASMANIAQAACLAFFLAKSEKL--KGLAGASGVSAVLGITPEP 387
Db 337 HIEMINGSGATYLLPIAAMAGAGQVGAALALWVRCKRNTTLRNTLKGALPV-GFLGIGEP 395
QY 388 AIFGVNRLRPWFI-----GIGTAATGGALIALFDIKAVAGAGFLGVSDIDAPDMVM 442
Db 396 LIYGVTLPLGRPPLTACIGGIGGAVIGG-----IGHIGAKAIGPSG-VSLPLISDNMYL 450
QY 443 FLVCANVTFFIAGAAATAYGL---YLVRNGSIDPDNATAAPVAGTTKA 488
Db 451 GYTAGLL-----AAYAGGFVCTYLF-----GTTKA 475

RESULT 7

US-09-815-242-5660
; Sequence 5660, Application US/09815242
; Patent No. US2002061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5660
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5660

Query Match

14.7%; Score 490.5; DB 9; Length 439;

Best Local Similarity 30.3%; Pred. No. 9.5e-36;
Matches 131; Conservative 71; Mismatches 166; Indels 65; Gaps 11;
QY 4 KDLAQRIIRDIGGDNIVAAAHACATRLRLVLKDTKDVDRQSLDDDDPLDKGTETGGMFQI 63
Db 5 QQLAERITAAVGCMNIDSVNMCMTVRKIVLDENKVKVDDQELRHIDGVNMGVIH-DERIQV 63
QY 64 IYGPQDGVHVFRELDATSKDIAVSTEQKDVVANNAN----- 101
Db 64 VVGPRTVKNVANHMAE-----LSGVKLGDTIPHNHNDSEKMDYKSYAADKAKANKEAH 116
QY 102 -----WFSRAVKVLADIPEVLPIILVG-----GGLLMAINNVVAQDLFGPOSLEMFPP 150
Db 117 KAKQNKGLNKVLKSIANIFILPIPAFIGAGLIGGIAAUNSLNLMVAGYISG-----A 168
QY 151 QISGVAEMINLMASAPFAELPVLVGFTATKRFSGNEFLGAGIGMAMVFPPLVNGYDVAAT 210
Db 169 WITQLITVFNVIKDMGLAYLAIFTGINAAKEFGATPGLGSGVIGGT-----TLLTG----- 218
QY 211 MTAGEMPWMSLGLDVAQAQYOGTVLPVLSVWILATIEKFLHKLMTADFLITPVLTL 270
Db 219 -IAGKNILMNVTGEPLQPG-QGGIIGVIFAVWILSIVERLKHIVPNAIDIIVTPTIAL 276
QY 271 LLTGFLTFIATGPAMRWGDLAHLGLOGLYDFGPGVGLLGLVYSPIVITGLHQSPPI 330
Db 277 LIVGLLTIETIEMPLAGFVSDLSVSVNGIISIGVFSGFIIGASFLPLVLMGLHHIFTPI 336
QY 331 ELEFNQ-QGSFIFATASMANIAQAACLAFFLAKSEKL--KGLAGASGVSAVLGITPEP 387
Db 337 HIEMINGSGATYLLPIAAMAGAGQVGAALALWVRCKRNTTLRNTLKGALPV-GFLGIGEP 395
QY 388 AIFGVNRLRPWFI-----GIGTAATGGALIALFDIKAVAGAGFLGVSDIDAPDMVM 442
Db 396 LIYGVTLPLGRPPLTACIGGIGGAVIGG-----IGHIGAKAIGPSG-VSLPLISDNMYL 450

RESULT 8

US-10-156-761-11459
; Sequence 11459, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11459
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11459

Query Match 14.1%; Score 472; DB 15; Length 545;

Best Local Similarity 25.9%; Pred. No. 6.3e-34;
Matches 152; Conservative 79; Mismatches 223; Indels 132; Gaps 14;
QY 3 HKDLAQRIIRDIGGDNIVAAAHACATRLRLVLKDTKDVDRQSLDDDDPLDKGTETGGMFQ 62
Db 6 HSNTAAAILPLYGGRANVTSAHCWTRLRLGLADSLYDEESLRALPAVLGVWD-DDTYQ 64
QY 63 IIVGPGDVHVFRELDATSKDIAVSTEQKDVVANNAN----- 97

Db 65 IVLPGIVAKVTPPEFKLLASSETTEAPANALPVYTAGAPTEAAPTAAEGPATSETPVAEP 124
QY 98 -----NNANWFSRAVKYVLADIFVPLIPILVGGGLMAI 130
Db 125 PARSAFEGTEAARLAARGAALKAAQORHATPVRLRLRIANIFVPLIPALIGCVIAGV 184
QY 131 NNVLVAQDLFGPQSLVEMFPOISGVAEMINLMASAPFAFLPVLVGFTATKRFSGNEPLGA 190
Db 185 NGLLNLGW-----LPSLTALAAV-----ASGFEMALIAVFGVHHTAKEFGTPIILGG 232
QY 191 GIGMAVFPPLVNGYDVAATMTAGEMPMWSLFGLDVAQAQYQGTVPVLYVWSILATIEK 250
Db 233 AVAAVIVYAGVAKVTFEGTILSPG-----QGVLCALAAALLGTIVVEK 275
QY 251 FLHRLMGTADFLTPVLTLLTGFLFI-----AIGPANWVWGDLAHLGLOGL 299
Db 276 WCRTWVPTDLVTPILTLLTGVLGLYGLMYAAGAAATSAIGTAATW-----L 324
QY 300 YDFGPGVGLLFGVLVSPIVITGLHQSFPPPIELEFNQGG-SFIFATASMANIAQAACL 358
Db 325 LSTGTGVEAGLTLGLFLPLVMLGLHQAFLPLHTLIEQOQYTVLLPVLAMAGAGQVGAAL 384
QY 359 AVEF-LAKSEKLGACAGSVSAVLGITEPAIGVNLRLRWPPYIGIGTAIGGALIALF 417
Db 385 AVIVLRHDSLRITIRSAIPAGLGVGEPLIYGVSLPLGRPELTACAGGAAGAFVGLF 444
QY 418 D-----IKAVALGAAGFL-----GVVSDIDAPDMVMFLVCAVTFVIAFGAAIAYGL 463
Db 445 SMLGDRIGSTAIGSPGWALPPLLAGSGLL-----PSAAIYAGLLTGAYVGFCAITYFGS 500
QY 464 YLV-----RRNSIDPDATAAPVPAGTTKAEAEAPAEFSNDSTI 502
Db 501 FGUVGVGCEAERTDIAAPTGTAAAPMT-TTPTDTAAPHDPDEGNRTL 545

RESULT 9

US-09-815-242-10243
; Sequence 10243, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10243
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-815-242-10243

Query Match 13.1%; Score 438; DB 9; Length 474;
Best Local Similarity 27.8%; Pred. No. 6e-31;
Matches 139; Conservative 84; Mismatches 213; Indels 64; Gaps 17;

QY 5 DIAQRLTRDIGEDNIVAAHCASTRRLVLKDFKVDQSLDDDDPDLLKSTETGGMFOII 64
Db 8 ELINTLTTRVGGPGNIASCNCMTLRUGVHDSIVD-PNITKLGKGVILTSQVQV 66
QY 65 VGPQGDVHDVFEK-----LDDATSKDIA-VSTEOLKDVVANNANWFSRAVKVLADIFVPLIP 119
Db 67 FPGKRAHRAAKAMSELLGEAPVQDAETAQNKROLKAKQTSVQOFLAKPATIFTPLIP 126
QY 120 ILVGGGLLMAINNVLVAQDLFGPQSLVEMFPOISGVAEMINLMASAPFAFLPVLVGFTAT 179
Db 127 GFIAAGLLGI-ATLIATVMHVPAQAQGTLPD---ALNFMKVFSGKLTFTFLVILVYNAA 182
QY 180 KRFGGNEFLGAGIGAMVFPPTLVNGYDVAATMTAGEMPMWSLFGLDVAQAQYQGTVPVLPV 239
Db 183 QAFG-----GTGVNGAIIAALFLLGYNPAAT-TGYAGFHDFFGLPIDP---RGNIIIGVL 233
QY 240 VVSWILATIEKFLHRLMGTADFLTPVLTLLTGFLFIAIGCPAMRWGDLAHLGLOGL 299
Db 234 IAAWACARTEGMVRRFMPDDLEMLTSLITLTITATLAYLIIMPLGGM-----LFEQMSWL 289
QY 300 YDF--GPPYVGGLLFGVLVSPIVITGLHQSFPPPIELELF-NOGGSIFATASMANIAQA 356
Db 290 FMHLNSNPGCAVLAGLFLIAVFGVHQGFIPVYLAALMDSQGFNSLFTLSWAGAGVGA 349
QY 357 CIAVFPFLAKSEK-LKGLAGASVSAVLGITEPAIGVNLRLRWPPYIGIGTAIGGALIA 415
Db 350 ALALYWRAPHSALRSQVRGAIIPGLLVGEPLIYGVTLPRMKPFV---TACLGGAAGG 405
QY 416 LF-----DRAVAL--GAAGFLGVVSDIDAPDMVMFLVCAVTFVIA 453
Db 406 LFGLIAWGLPMGLNSAPGSPGLVALPLMTSAQGLPAMAVYAGIILVAVVCGFI-FTT 464
QY 454 AFCAAIAYGLYLVRNGSID 473
Db 465 LFQC-----RNVNLD 474

RESULT 10

US-09-815-242-5658
; Sequence 5658, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

```

: APPLICANT: Carr, Grant J
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: PROKARYOTES
: FILE REFERENCE: ELIURA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21

```

; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR FILING DATE: 2000-05-23

;
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/252 626

; PRIOR APPLICATION NUMBER: 00/233,023
 ; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
 ;
 ; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/269,308

```

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12270

```

```

; LENGTH: 681

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus aureus

```

US-09-815-242-12270

Query Match	11.0%	Score 368.5	DB 9	Length 681:
-------------	-------	-------------	------	-------------

Best Local Similarity	23.3%	Pred. No. 2.1e-24;
Matches 160: Conservative	85:	Mismatches 241:
	Indels	201: Gaps
		23:

[illegible]

QY 103 FSRVAVLADIEVPLIPILVGGGLMAINNVLAQDL--FGPQSLVEMFPQISGVAEMIN 180

Db 6 FGQLRIGKALMLP-VAILPAAGLLAIGTAIQGEALQHYP-----FIQNGGVQNVAK 58

QY 161 LMASAP---FAFLPVLVGTTATKRFEGGNEFLGAGIGMAMVFPTLVNGYDVAA----- 209

```
Db      |||::|||:|||||
59 LMTAAGSIIFFENLPMIFA-----LGVAIGLA-----GGDGVAAIAAFVGYI 99
```

[illegible]

QY ZIU -IMAGEM P-----PMWSLFGLDVAQAQIGIVLFLVLSWILPATIENT LBN 2304
| | : | : | | | : : | | :
| | : | : | | | : : | | :

DB 100 IMNKTMGDFLQVTPKNVTD PASGYASILGIP TLQTGVFGGIIIGALAAW--CYNKFYNI 156

QY 255 RLMGTADEL---ITPVLTLTLTGELTFIAIGPAMRWVGDLLAHLQ---GLYDFGGPV 306

D_b 157 NLPSYLGFFAGKRFVPIM-MATTSFI--LAFPMALIW--PTIQTGLNAFSTGLLDSNTGV 211

307 COLLECTIVIST HOSE
330 DDT BT BT ENOC

[illegible]

Db Db 212 AVFLFGFIKRLLPFGLHHIFHAPFWFEFGSWKNAAGEIIHGDQRI FIEQIREGAHLTAG 271

QY 340 SFIFATASMANIAQGAACLA VFFELAKSEKLKGLAGASVSA ---VLGITEPAIFGVNLR 395

Db 272 KFMQGEFFVMFGLPAAALAIYHSAKPENKKVWAGLMGSAALTSFLTGITTEPLEFSFLFV 331

OW 306 I BWDEVTCTCTEANTICCAI IAT EDTKAVAVI CARCEI CIVIXSTDA - DDMMME-----I VCA 447

QY	390	395	400	405	410	415	420	425	430	435	440	445	450	455	460	465	470	475	480	485	490	495	500	505	510	515	520	525	530	535	540	545	550	555	560	565	570	575	580	585	590	595	600	605	610	615	620	625	630	635	640	645	650	655	660	665	670	675	680	685	690	695	700	705	710	715	720	725	730	735	740	745	750	755	760	765	770	775	780	785	790	795	800	805	810	815	820	825	830	835	840	845	850	855	860	865	870	875	880	885	890	895	900	905	910	915	920	925	930	935	940	945	950	955	960	965	970	975	980	985	990	995	1000
QY	390	395	400	405	410	415	420	425	430	435	440	445	450	455	460	465	470	475	480	485	490	495	500	505	510	515	520	525	530	535	540	545	550	555	560	565	570	575	580	585	590	595	600	605	610	615	620	625	630	635	640	645	650	655	660	665	670	675	680	685	690	695	700	705	710	715	720	725	730	735	740	745	750	755	760	765	770	775	780	785	790	795	800	805	810	815	820	825	830	835	840	845	850	855	860	865	870	875	880	885	890	895	900	905	910	915	920	925	930	935	940	945	950	955	960	965	970	975	980	985	990	995	1000

Db 332 AP L L F F I H A V L D G L S F L T L Y L L D V H L G Y T F S G G F I D Y L L G V L P N K T Q W L V I P V G L V Y A 391

QY 448 VTF-----VIAFGAAIAYGLY-----LVR- 467

```

Db      392 V I Y Y F V F R E L I V K L K Y K T P G R E D K Q S Q A V T A S A T E L P Y A V L E A M G G K A N I K H L D A C I T R L 451

```

168 -----PNCSTDDDTAA 479

1	:	:	
2	:	:	
3	:	:	
4	:	:	
5	:	:	
6	:	:	
7	:	:	
8	:	:	
9	:	:	
10	:	:	
11	:	:	
12	:	:	
13	:	:	
14	:	:	
15	:	:	
16	:	:	
17	:	:	
18	:	:	
19	:	:	
20	:	:	
21	:	:	
22	:	:	
23	:	:	
24	:	:	
25	:	:	
26	:	:	
27	:	:	
28	:	:	
29	:	:	
30	:	:	
31	:	:	
32	:	:	
33	:	:	
34	:	:	
35	:	:	
36	:	:	
37	:	:	
38	:	:	
39	:	:	
40	:	:	
41	:	:	
42	:	:	
43	:	:	
44	:	:	
45	:	:	
46	:	:	
47	:	:	
48	:	:	
49	:	:	
50	:	:	
51	:	:	
52	:	:	
53	:	:	
54	:	:	
55	:	:	
56	:	:	
57	:	:	
58	:	:	
59	:	:	
60	:	:	
61	:	:	
62	:	:	
63	:	:	
64	:	:	
65	:	:	
66	:	:	
67	:	:	
68	:	:	
69	:	:	
70	:	:	
71	:	:	
72	:	:	
73	:	:	
74	:	:	
75	:	:	
76	:	:	
77	:	:	
78	:	:	
79	:	:	
80	:	:	
81	:	:	
82	:	:	
83	:	:	
84	:	:	
85	:	:	
86	:	:	
87	:	:	
88	:	:	
89	:	:	
90	:	:	
91	:	:	
92	:	:	
93	:	:	
94	:	:	
95	:	:	
96	:	:	
97	:	:	
98	:	:	
99	:	:	
100	:	:	

Db 452 RVEVNDKSKVDVPGLKDLGASGVLEVGNMMAIFGPKSDQIKHEMQQIMNGQVVENPTM 511

QY 480 PVPAGTTKAEAEAPAEFSNDSTIIQAPLTGEAIALSSVSDAMFASGKLGSGVAIVPTKGQ 539

```
Db      512 EDDKDET V V V A E D K S A T S E L S H I V H A P L T G E V T P L S E V P D Q V F S E K M M G D G I A I K P S Q G E 571
```



```
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10089
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10089

Query Match      9.23; Score 308.5; DB 9; Length 648;
Best Local Similarity 22.18; Pred. No. 5.2e-19;
Matches 149; Conservative 85; Mismatches 241; Indels 199; Gaps 21;

QY 152 ISGVAEMINLMASAFFFLVYGVFTATKFGGNEFLG-AGIGMA--MVFPFLVNGYDVA 208
DQ 152 ISGVAEMINLMASAFFFLVYGVFTATKFGGNEFLG-AGIGMA--MVFPFLVNGYDVA 208
DB 3 ILGFFQRLGRALQPLPIAVLPVA---ALLRFQGPDLNVAFTAQAGGAIFDNL-----A 53
QY 209 ATMTAGEMPWMSLGLDVAOAGYQGTVPVTVVSWTILATIEKFLKRLMGTADEFLITPVL 268
DQ 209 ATMTAGEMPWMSLGLDVAOAGYQGTVPVTVVSWTILATIEKFLKRLMGTADEFLITPVL 268
DB 54 LIFAIGVASSHS---KDSRAAALAGAVGYFLTKAMVTINPEIN---MGVLAGIITGLV 107
QY 269 -----TLLFGFTTF-----IAIGPAMRWVDLLAHGLOGLYDF-- 302
DQ 269 -----TLLFGFTTF-----IAIGPAMRWVDLLAHGLOGLYDF-- 302
DB 108 GGAAYNRWSDIKLPDFLSFTFGKRFVPIATGFFCLVLAIFGYVWPVQVHAHAGGEIV 167
QY 303 -GGPVGGLLGLVSPVITVIGLHQSPPPI-----ELEFNQO---G 339
DQ 303 -GGPVGGLLGLVSPVITVIGLHQSPPPI-----ELEFNQO---G 339
DB 168 SAGALSGSIFGFIRNLLIPTGLHQLVNTIAWFOIGFTNNAAGVVFHGDINRFYAGDGTAG 227
QY 340 SFIFATASMANIAGACLAFLAKSEK---LKGLAGASGVSAVL-GITEPAIF----- 390
DQ 340 SFIFATASMANIAGACLAFLAKSEK---LKGLAGASGVSAVL-GITEPAIF----- 390
DB 228 MEMSGFFPIMFGLPGGAALAMYFAAPKXERPMVGMGLLSVAVTAFTLTGVTPELFLEMF 287
QY 391 -----GVNLRWRPFYIGTAAIGGAL----- 413
DQ 391 -----GVNLRWRPFYIGTAAIGGAL----- 413
DB 288 APLLYLLHALTGISLFAVTLIGTHAGFSAGAGIDYALMYNLPAASQNVWMLVMGVIF 347
QY 414 -----IALFDIKA-----VALGAA----- 427
DQ 414 -----IALFDIKA-----VALGAA----- 427
DB 348 FAIYFVFSVLIRMFENLKTGREKDEIVTEFANSNTEEGLFQLATNYIAAVGGTDNLK 407
QY 428 -----GELGVYSIDAPDMVMPFLVCVAVTFVIAFGAAIA 460
DQ 428 -----GELGVYSIDAPDMVMPFLVCVAVTFVIAFGAAIA 460
DB 406 AIDACITELRLTVADSRVNDTCKRLGASGVYKLNKQTQIVIVGAKAESIGDAMKKVVA 467
QY 461 YGLVLRRENGSIDPDAAPVAGTTRKAERAEAPAFSNDSTIIQAPLTGEATIALSSVSDA 520
DQ 461 YGLVLRRENGSIDPDAAPVAGTTRKAERAEAPAFSNDSTIIQAPLTGEATIALSSVSDA 520
DB 468 RG-PYAAASAEATP-ATAAPVAKPQAVPNVASTAE-----LVSPITGDVVALDQVPDE 518
QY 521 MFASGKLGSGVAIVPTGQLVSPVSGKIVVAFPSGHAFVTRKAEDGSNNVDILMHIGFDT 580
DQ 521 MFASGKLGSGVAIVPTGQLVSPVSGKIVVAFPSGHAFVTRKAEDGSNNVDILMHIGFDT 580
DB 519 AFASKAVGCVAVKPTCKIVVSPAAGTIVKIFNTNHAFCLETE---KGAEIVVHNGIDT 574
QY 581 VNLNTHNPNLKKOGDEKAGELCEFDIDAIRKAAGYEVTTPIVWSNKKTKGPNVTYGLG 640
DQ 581 VNLNTHNPNLKKOGDEKAGELCEFDIDAIRKAAGYEVTTPIVWSNKKTKGPNVTYGLG 640
DB 575 VALEGKGFKRLVEGAQVSAQPILEMDLDLYLNARSMISPVVCSNIDDFSGLIKAQG 634
QY 641 EIEAGANLLNVAKK 654
```

Db 635 HIVAGQTPLYEIKK 648

Search completed: October 1, 2003, 05:15:36
Job time : 220 secs